Peter Schuster

Stochasticity in Processes

Fundamentals and applications to chemistry and biology

October 27, 2014

Springer

Preface

Statistics and stochastic processes are often neglected in the education of chemists and biologists, although modern experimental techniques allow for investigations of small sample sizes down to single molecules and many measured data are sufficiently accurate for direct detection of fluctuations. Progress in the development of new techniques and improvement in the resolution of conventional experiments have been enormous within the last fifty years. Indeed, molecular spectroscopy provided hitherto unimaginable insights into processes down to range of hundred attoseconds with atomic resolution, observations on single particles became routine and current theory in physics, chemistry, and the life sciences cannot be successful without a deeper understanding of randomness and its causes. Sampling of data and reproduction of processes are doomed to produce artifacts in the interpretation unless the observer has a solid background in the mathematics of limited reproducibility. As a matter of fact stochastic processes are much closer to observations than deterministic descriptions in modern science and everyday life. Exceptions are, of course, the motions of planets and moons as encapsulated in celestial mechanics, which stood at the beginnings of science and modeling by means of differential equations. Fluctuations are so small that they cannot be detected, even not by highest precision measurements: Sunrise, sunset, and solar eclipses are predictable with practically no scatter. Processes in the life sciences are often entirely different. A famous and typical historical example are Mendel's laws of inheritance: Regularities are detectable only in sufficiently large samples of individual observations, and the influence of stochasticity is ubiquitous. Processes in chemistry are between the two extremes: The deterministic approach in conventional chemical reaction kinetics has neither suffered a loss in applicability nor did the results become less reliable in the light of modern experiments. What has increased rather dramatically are the accessible resolutions in detectable amounts of materials in both, space and time. Deeper insights into mechanisms provide new access to information on molecular properties for theory and practice.

Biology is currently in a state of transition: The molecular connection to chemistry revolutionized the sources of biological data and is setting the stage for a new theoretical biology. Historically, biology was based almost exclusively on observation, and theory in biology was engaged only in the interpretations of observed regularities. The development of biochemistry at the end of the nineteenth and the first half of twentieth century introduced quantitative thinking in terms of chemical kinetics into some biological subdisciplines. Biochemistry attributed also a new dimension to experiments in biology in the form of *in vitro* studies on isolated and purified biomolecules. A second import of mathematics into biology came in the form of population genetics, which was created in the nineteen twenties as a new theoretical discipline uniting Darwin's natural selection and Mendelian genetics. This happened in the theoretical approach more than twenty years before evolutionary biologists completed the so-called *synthetic theory* performing the same goal. Beginning in the second half of the twentieth century molecular biology started to build a solid bridge from chemistry to biology and enormous progress in experimental techniques created a previously unknown situation in biology insofar as new procedures were required for data handling, analysis, and interpretation since the volume of information is drastically exceeding the capacities of human mind. Biological cells and whole organisms become now accessible to complete description at the molecular level. The overwhelming amount of information that is required for a deeper understanding of biological objects is a consequence of two factors: (i) the complexity of biological entities and (ii) the lack of a universal theoretical biology.

The current flood of results from molecular genetics and genomics to systems biology and synthetic biology requires – apart from elaborate computer techniques – primarily suitable statistical methods and tools for verification and evaluation of data. Analysis, interpretation, and understanding of experimental results, however, is impossible without proper modeling tools. In the past these tools were primarily based on differential equations but it has been realized within the last two decades that an extension of the available repertoire by stochastic methods inevitable. Moreover, the enormous complexity of the genetic and metabolic networks in the cell calls for radically new methods of modeling that resemble the mesoscopic level of description in solid state physics. In mesoscopic models the overwhelming and for many purposes dispensable wealth of detailed molecular information is cast into a partially probabilistic description in the spirit of *dissipative particle dynamics*, and such a description cannot be successful without a solid mathematical background. The field of stochastic processes has not been bypassed by the digital revolution. Numerical calculation and computer simulation play a decisive role in present day stochastic modeling in physics, chemistry and biology. Speed of computation and digital storage capacities are growing exponentially since the nineteen sixties with an approximate doubling time of eighteen month, a fact that is commonly addressed as Moore's law [338]. It is not so well known, however, that the spectacular exponential growth in

computer power has been overshadowed by the progress in numerical methods that led to an enormous increase in the efficiency of algorithms. To give just one example, which was reported by Martin Grötschel from the Konrad Zuse-Zentrum in Berlin [215, p. 71]: The solution of a benchmark production planning model by linear programming would have taken – extrapolated – 82 years CPU time in 1988, using the computers and the linear programming algorithms of the day. In 2003 – fifteen years later – the same model could be solved in one minute and this means an improvement by a factor of about 43 million. Out of this, a factor of roughly 1 000 resulted from the increase in processor speed whereas a factor of 43 000 was due to improvement in the algorithms. Many other examples of similar progress in the design of algorithms can be given. Understanding, analyzing, and designing of high-performance numerical methods, however, requires a firm background in mathematics. The availability of cheap computing power has also changed the attitude towards exact results in terms of complicated functions: It does not take so much more computer time to compute a sophisticated hypergeometric function than to calculate an ordinary trigonometric function for an arbitrary argument, and operations on confusingly complicated expressions are enormously facilitated by symbolic computation. In this way the present day computational facilities have also large impact on the analytical work.

In the past biologists had often mixed feelings for mathematics and reservations against the use of too much theory. The new developments, however, have changed the scene: The enormous amount of data that are collected by the new techniques can neither be inspected by human eyes nor comprehended by human brains, sophisticated software is required for handling and analysis, and modern biologists have to rely on it [394]. The biologist Sydney Brenner, an early pioneer of molecular life sciences, points out [52]: *"... But of course we see the most clear-cut dichotomy between hunters and gatherers in the practice of modern biological research. I was taught in the pregenomic era to be a hunter. I learnt how to identify the wild beasts and how to go out, hunt them down and kill them. We are now, however, being urged to be gatherers, to collect everything lying about and put it into storehouses. Someday, it is assumed, someone will come and sort through the storehouses, discard all the junk and keep the rare finds. The only difficulty is how to recognize them."* The recent developments in molecular biology, genomics and organismic biology, however, seem to initiate this change in biological thinking since there is practically no chance to shape modern life sciences without mathematics, computer science and theory. Brenner advocates the development of a comprehensive theory that would provide the proper frame for modern biology [51]. He and others are calling for a *theoretical biology new* that allows for handling the enormous biological complexity. Manfred Eigen stated very clearly what we can expected from such a theory [91, p. xii]: *"Theory cannot remove complexity but it can show what kind of 'regular' behavior can be expected and what experiments have to be done to get a grasp on the irregularities."* Among the things, the new theoretical biology will have to find

the appropriate way to combine randomness and deterministic behavior in modeling and it is not very risky to predict that it will need a strong anchor in mathematics in order to be successful.

In this monograph an attempt is made to collect the necessary mathematical background material for understanding stochastic processes and its applications in chemistry and biology. In the sense of Albert Einstein's version of Occam's razor [56, pp. 384-385; p. 475], *"... Everything should be made as simple as possible, but not simpler. ..."*, dispensable deep dwelling in higher mathematics has been been avoided. In particular, an attempt was made to keep the requirements in mathematics at the level of an undergraduate mathematics course for scientists, and we tried to make the monograph as much self-contained as possible. Derivations of key equations are given wherever this can be done with reasonable mathematical efforts. The derivations of analytical solutions for selected examples are given in full length because the reader who is interested to apply the theory of stochastic processes in practice should be brought in the position to derive new solutions on his own. Some sections that are not required if one is primarily interested in applications are marked for skipping by readers who are willing to accept the basic results without explanations. The book is partitioned into six chapters: The first chapter provides an introduction into probability theory and follows in part the concept of the introduction into probability theory by Kai Lai Chung [69], chapter two deals with the link between abstract probabilities and measurable quantities through statistics, chapter three describes stochastic processes and their analysis, and has been partly inspired by a monograph by Crispin Gardiner [157], the following chapters four and five present selected applications of stochastic processes to solving problems in chemistry and biology and the closing chapter contains a brief outlook on expected future developments. Throughout the book the focus is laid on stochastic methods and the origin of in the various equation is not discussed with one exception: chemical kinetics. For this example we present two sections on theory and empirical determination of reaction rate parameters, because in this case it is possible to show how Ariadne's red thread guides from first principles in theoretical physics to the equations of stochastic chemical kinetics. We refrained from preparing a separate section with exercises, instead case studies, which may serve as good examples for calculations done by the reader himself, are indicated in the book. Sources from literature were among others the text books [69, 115, 157, 161, 174, 213, 301, 435, 463]. For a brief and concise introduction we recommend [231]. Standard textbooks in mathematics used for the courses were: [14, 46, 317, 382]. For dynamical systems theory the monographs [183, 209, 404, 418] are recommended.

This book is derived from the manuscript of a course in stochastic chemical kinetics for graduate students of chemistry and biology held in the years 1999, 2006, 2011, and 2013. Comments by the students of all four courses were very helpful in the preparation of this text and are gratefully acknowledged. Several colleagues gave important advice and critically read the manuscript,

among them Edem Arslan, Christoph Flamm, Thomas Hoffmann-Ostenhof, Christian Höner zu Siederissen, Ian Laurenzi, Eberhard Neumann, Paul E. Phillipson, Karl Sigmund, and Peter F. Stadler. Many thanks to all of them.

Wien, *Peter Schuster*
October 2014

Contents

Chapter 1
Probability

Who considers too much will achieve little.
Wer gar zu viel bedenkt, wird wenig leisten.
Friedrich Schiller, Wilhelm Tell, III.

Abstract . Thinking in terms of probability originated historically from analyzing the chances of success in gambling and its mathematical foundations were laid down together with the development of statistics in the seventeenth century. Since the beginning of the twentieth century statistics is an indispensable tool for bridging the gap between molecular motions and macroscopic observations. The classical notion of probability is based on counting and dealing with finite numbers of observations, the extrapolation to limiting values for hypothetical infinite numbers of observations is the basis of the frequentists' interpretation, and more recently a *subjective* approach derived from the early works of Bayes became useful in modeling and analyzing complex biological systems. The Bayesian interpretation of probability accounts explicitly for incomplete and improvable knowledge of the experimenter. In the twentieth century set theory became the ultimate basis of mathematics and in this sense it became also the fundament of current probability theory that is based on Kolmogorov's axiomatization in 1933. The modern approach allows for handling and comparing countable, countable infinite and the most important class of uncountable sets, which are underlying continuous variables. Borel fields being uncountable subsets of sample spaces allow for defining probabilities for certain uncountable sets like, for example, the real numbers. The notion of random variables is central to the analysis of probabilities and applications to problem solving. Random variables are characterized conventionally in form of their distributions in discrete and countable or continuous and uncountable probability spaces.

Classical probability theory, in essence, can handle all cases that are modeled by discrete quantities. It is based on counting and accordingly runs into problems when it is applied to uncountable sets. Uncountable sets, however, occur with continuous variables and are indispensable therefore for modeling processes in space as well as for handling large particle numbers, which are described in terms of concentrations in chemical kinetics. Current probability theory is based on set theory and can handle variables on discrete – and

countable – as well as continuous – and uncountable – sets. After a general introduction we present historical probability theory by means of examples, different notions of probability are compared, and then we provide a short account of probabilities, which are axiomatically derived from set theoretical operations. Separate sections are dealing with countable and uncountable sample spaces. Random variables are characterized in terms of probability distributions and their properties will be introduced and analyzed insofar as they will be required in the applications to stochastic processes.

1.1 Fluctuations and precision limits

An scientist reproduces an experiment. What is he expecting to observe? If he were a physicist of the early nineteenth century he would expect the same results within the precision limits of the apparatus he is using for the measurement. Uncertainty in observations was considered to be merely a consequence of technical imperfection. Celestial mechanics comes close to this ideal and many of us, for example, could witness the enormous accuracy of astronomical predictions in the precise dating of the eclipse of the sun in Europe on August 11, 1999. Terrestrial reality, however, tells that there are limits to reproducibility that have nothing to do with lack of experimental perfection. Uncontrollable variations in initial and environmental conditions on one hand and large intrinsic diversity of the individuals in a population on the other hand are daily problems in biology. Limitations of correct predictions are commonplace in complex systems: We witness them every day by watching the failures of various forecasts from the weather to the stock market. Another not less important source of randomness comes from irregular thermal motions of atoms and molecules that are commonly characterized as thermal fluctuations. The importance of fluctuations in the description of ensembles depends on the population size: They are – apart from exceptions – of moderate importance in chemical reaction kinetics but highly relevant for the evolution of populations in biology.

Conventional chemical kinetics is handling ensembles of molecules with large numbers of particles, $N \approx 10^{20}$ and more. Under the majority of common conditions, for example near or at chemical equilibria or stable stationary states and in absence of autocatalytic self-enhancement, random fluctuations in particle numbers are proportional to $\sqrt{N}$. Dealing with substance amounts of about 10^{-4} moles – being tantamount to $N = 10^{20}$ particles – natural fluctuations involve typically $\sqrt{N} = 10^{10}$ particles and thus are in the range of $\pm 10^{-10} N$. Under these conditions the detection of fluctuations would require a precision in the order of $1 : 10^{10}$, which is (almost always) impossible to achieve.[1] Accordingly, the chemist uses concentrations rather than particle numbers, $c = N/(N_\mathrm{L} \cdot V)$ wherein $N_\mathrm{L} = 6.23 \times 10^{23}$ and V are Avogadro's number[2] and the volume in dm^3, respectively. Conventional chemical kinetics considers concentrations as continuous variables and applies deterministic

[1] Most techniques of analytical chemistry meet serious difficulties when accuracies in concentrations of 10^{-6} or higher are required.

[2] The amount of a chemical compound **A** is commonly measured as the number of molecules, N_A, in the reaction volume V or as concentrations c_A being the numbers of moles in one liter of solution, $c_\mathsf{A} = N_\mathsf{A}/(N_\mathrm{L}\, V)$ where N_L is Avogadro's constant, which is closely related to Loschmidt's number. The difference between the two quantities that is often ignored in the literature: Avogadro's number, $N_\mathrm{L} = 6.02214179 \times 10^{23}\,\mathrm{mol}^{-1}$ refers to one mole substance whereas Loschmidt's constant $n_0 = 2.6867774 \times 10^{25}\,\mathrm{m}^{-3}$ counts the number of particles in one liter gas under normal conditions. The conversion factor between both constants is the molar volume of an ideal gas that amounts to $22.414\ \mathrm{dm}^3 \cdot \mathrm{mol}^{-1}$.

methods, in essence differential equations, for modeling and analysis of reactions. Thereby, it is implicitly assumed that particle numbers are sufficiently large that the limit of infinite particle numbers neglecting fluctuations is correct. This scenario is commonly not fulfilled in biology where particle numbers are much smaller than in chemistry and uncontrollable environmental effects introduce additional uncertainties.

Nonlinearities in chemical kinetics may amplify fluctuations through autocatalysis and then the random component becomes much more important than the $\sqrt{N}$-law suggests. This is the case already with simple autocatalysis discussed in section 5.1 and becomes a dominant effect, for example, with oscillating concentrations and deterministic chaos. Some processes in physics, chemistry, and biology have no deterministic component at all, the most famous of it is Brownian motion, *Brownian motion* which can be understood as a visualized form of diffusion. In biology other forms of entirely random processes are encountered as well where fluctuations are the only or the major driving force of change. An important example is random drift of population in the space of genotypes in absence of fitness differences or fixation of mutants in evolution where each new molecular species starts out from a single variant.

In 1827 the British botanist Robert Brown detected and analyzed irregular motions of particles in aqueous suspensions that turned out to be independent of the nature of the suspended materials – pollen grains, fine particles of glass or minerals served equally well [55]. Although Brown himself had already demonstrated that the motion is not caused by some (mysterious) biological effect, its origin remained kind of a riddle until Albert Einstein [108], and independently Marian von Smoluchowski [453], published satisfactory explanations in 1905 and 1906,[3] which revealed two main points:

(i) The motion is caused by highly frequent collisions between the pollen grain and the steadily moving molecules in the liquid in which it is suspended, and
(ii) the motion of the molecules in the liquid is so complicated and irregular that its effect on the pollen grain can only be described probabilistically in terms of frequent, statistically independent impacts.

In order to model Brownian motion Einstein considered the number of particles per volume as a function of space and time, $f(x,t) = N(x,t)/V$,[4] and derived the equation

[3] The first mathematical model of Brownian motion has been conceived already in 1880 by Thorvald Thiele [274, 428]. Later in 1900 a process using random fluctuations of the Brownian motion type was used by Louis Bachelier [24] in order to describe the stock exchange market at the bourse in Paris. He gets the credit for having been the first to write down and analyze a Langevin equation (section 3.4). For a recent and detailed excellent monograph on Brownian motion and the mathematics of normal diffusion we recommend [174].

[4] For the sake of simplicity we consider only motion in one spatial direction, x.

$$\frac{\partial f}{\partial t} = D\frac{\partial^2 f}{\partial x^2} \quad \text{with the solution} \quad f(x,t) = \frac{\varrho}{\sqrt{4\pi D}}\,\frac{\exp\left(-x^2/(4Dt)\right)}{\sqrt{t}}\ ,$$

where $\varrho = N/V = \int f(x,t)\,\mathrm{d}x$ is the total number of particles per unit volume and D is a parameter called the *diffusion coefficient*. Einstein showed that his equation for $f(x,t)$ is identical with the differential equation of diffusion already known as Fick's second law [136], which had been derived fifty years earlier by the German physiologist Adolf Fick. Einstein's original treatment is based on small discrete time steps $\Delta t = \tau$ and thus contains a – well justifiable – approximation that can be avoided by application of the current theory of stochastic processes (section 3.2.2.2). Nevertheless Einstein's publication [108] represents the first analysis based on a probabilistic concept that is by all means comparable to the current theories and Einstein's paper is correctly considered as the beginning of stochastic modeling. Later Einstein wrote four more papers on diffusion with different derivations of the diffusion equation [109]. It is worth mentioning that three years after the publication of Einstein's first paper Paul Langevin presented an alternative mathematical treatment of random motion [271] that we shall discuss at length in the form of the Langevin equation in section 3.4. Since the days of Brown's discovery the interest in Brownian motion has never ceased and publications on recent theoretical and experimental advances document this fact nicely, two interesting recent examples are [284, 399].

Einstein computed from the solution of the diffusion equation that the diffusion parameter D is linked to the mean square displacement in the x-direction, which the particle undergoes during time t – $\langle \Delta x^2 \rangle$ or its square root λ_x:

$$D = \frac{\langle \Delta x^2 \rangle}{2t} \quad \text{and} \quad \lambda_x = \sqrt{\langle \Delta x^2 \rangle} = \sqrt{2Dt}\ .$$

Extension to three dimensional space is straightforward and results only in a different numerical factor: $D = \langle \Delta x^2 \rangle / (6t)$. Both quantities, the diffusion parameter D and the mean displacement λ_x are measurable and Einstein concluded correctly that a comparison of both quantities should allow for an experimental determination of Avogadro's number [367].

Brownian motion was indeed the first completely random process that became accessible to a description within the standards of classical physics. Previously, thermal motion had been identified as the driving force causing irregular collisions of molecules in gases by James Clerk Maxwell and Ludwig Boltzmann but the physicists in the second half of the nineteenth century were not interested in the details of molecular motion unless they were required in order to describe systems in the thermodynamic limit. In statistical mechanics the measurable macroscopic functions were and are derived by means of global averaging techniques. Thermal motion as an uncontrollable source of random natural fluctuations has been supplemented in the first

half of the twentieth century by quantum mechanical uncertainty as another limitation of achievable precision.

The occurrence of complex dynamics in physics and chemistry has been known since the beginning of the twentieth century through the pathbreaking theoretical works of the French mathematician Henri Poincaré and the experiments of the German chemist Wilhelm Ostwald who explored chemical systems showing periodicities in space and time. Systematic studies of dynamical complexity, however, required assistance by electronic computers and the new field of research on complex dynamical systems was initiated not before the nineteen sixties. The first pioneer of this discipline was Edward Lorenz [294] who detected what is nowadays called *deterministic chaos* through numerical integration of differential equations. New in the second half of the twentieth century were not so much the ideas but the tools to study complex dynamics. Quite unexpectedly, easy access to previously unknown computer power and the development of highly efficient algorithms made numerical computation to an indispensable source of scientific information that by now became almost equivalent to theory and experiment. Computer simulations have shown that a large class of dynamical systems modeled by nonlinear differential equations exhibits irregular – that means nonperiodic – variations for certain ranges of parameter values. Limited predictability of complex dynamics is highly important in practice: Although the differential equations used to describe and analyze chaos are still deterministic, initial conditions of a precision that can never be achieved in reality would be required for correct longtime predictions. Sensitivity to small changes makes a stochastic treatment indispensable, and solutions were found to be extremely sensitive to small changes in initial and boundary conditions in these *chaotic regimes*. Solution curves that are almost identical at the beginning, deviate exponentially from each other and appear completely different after sufficiently long time. Deterministic chaos gives rise to a third kind of uncertainty, because initial conditions cannot be controlled with higher precision than the experimental setup allows. It is not accidental that Lorenz detected chaotic dynamics first in the equations for atmospheric motions, which are indeed so complex that forecast is confined to short and medium time spans. Here we shall focus on the mathematical handling of processes that are irregular and often simultaneously sensitive to small changes in initial and environmental conditions, and we shall not be concerned with the physical origin of these irregularities.

1.2 The history of thinking in terms of probability

The concept of probability originated much earlier than its applications in physics and resulted from the desire to analyze gambling by rigorous mathematical methods. An early study that has largely remained unnoticed is due to the sixteenth century Italian mathematician Gerolamo Cardano contained already the basic ideas of probability. The beginning of classical probability theory is commonly associated with the story of French mathematician Blaise Pascal and the professional gambler, the Chevalier de Méré, which took place in France 100 years after Cardano. This narrative is such a nice illustration of a pitfall in applying probabilistic thinking that we repeat it here as our first example of conventional probability theory despite the fact that it is found in almost every textbook on statistics and probability.

In a letter of July 29, 1654, Blaise Pascal addressed a letter to the French mathematician Pierre de Fermat, which reports the careful observation of the professional gambler Chevalier de Méré who recognized that obtaining at least one *six* with one die in 4 throws is successful in more than 50% whereas obtaining at least two times the "six" with two dice in 24 throws has less than 50% chance to win. He considered this finding as a paradox, because he had calculated naïvely and erroneously that the chances should be the same:

$$\text{4 throws with one die yields} \quad 4 \times \frac{1}{6} = \frac{2}{3} ,$$

$$\text{24 throws with two dice yields} \quad 24 \times \frac{1}{36} = \frac{2}{3} .$$

Blaise Pascal became interested in the problem and calculated correctly the probability as we do it now in classical probability theory by counting of events:

$$\text{probability} \quad = \quad \text{Prob} \quad = \quad \frac{\text{number of favorable events}}{\text{total number of events}} . \qquad (1.1)$$

Probability according to equation (1.1) is always a positive quantity between zero and one, $0 \leq P \leq 1$. The sum of the probabilities that a given event has either occurred or did not occur thus is always one. Sometimes, as in Pascal's example, it is easier to calculate the probability of the unfavorable case, q, and to obtain the desired probability by computing $p = 1 - q$. In the one-die example the probability not to throw a *six* is 5/6, in the two-dice case the probability not to obtain two *six* is 35/36. Provided the events are independent their probabilities are multiplied[5] and we finally obtain for 4 and 24 trials, respectively:

[5] We shall come back to a precise definition of independent events later when we introduce current probability theory in section 1.6.4.

$$q(1) = \left(\frac{5}{6}\right)^4 \quad \text{and} \quad p(1) = 1 - \left(\frac{5}{6}\right)^4 = 0.5177 \,,$$

$$q(2) = \left(\frac{35}{36}\right)^{24} \quad \text{and} \quad p(2) = 1 - \left(\frac{35}{36}\right)^{24} = 0.4914 \,.$$

It is remarkable that Chevalier de Méré could observe this rather small difference in the probability of success – indeed, he must have watched the game very often!

The second example presented here is the *birthday problem*.[6] It can be used to demonstrate the common human weakness in estimating probabilities:

> *"Let your friends guess – without calculating – how many persons you need in a group such that there is a fifty percent chance that at least two of them celebrate their birthday on the same day. You will be surprised by the oddness of some of the answers!"*

With our knowledge on the gambling problem this probability is easy to calculate. First we compute the negative event: all persons celebrate their birthdays on different days in the year – 365 days, no leap-year – and find for n people in the group,[7]

$$q = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \ldots \cdot \frac{365-(n-1)}{365} \quad \text{and} \quad p = 1 - q \,.$$

The function $p(n)$ is shown in figure 1.1. For the above mentioned 50% chance we need only 27 persons, with 41 people we have already more than 90%

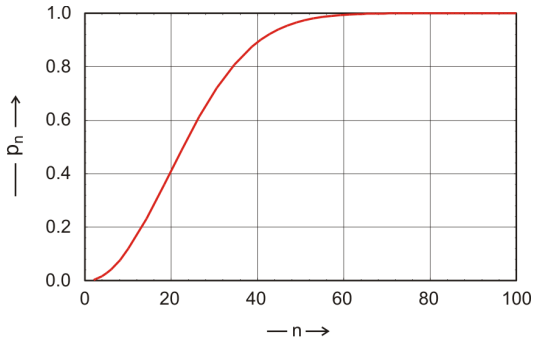


Fig. 1.1 The birthday problem. The curve shows the probability p_n that two persons in a group of n people celebrate birthday on the same day of the year.

[6] The birthday problem has been invented in 1939 by Richard von Mises [452] and it has fascinated mathematicians ever since. It was discussed and extended in many papers, for example [2, 74, 211, 351], and found its way into textbooks on probability theory [133, pp. 31-33].

[7] The expression is obtained by the argument that the first person can choose his birthday freely. The second person must not choose the same day and so he has 364 possible choices. For the third remain 363 choices and the nth person, ultimately, has $365 - (n-1)$ possibilities.

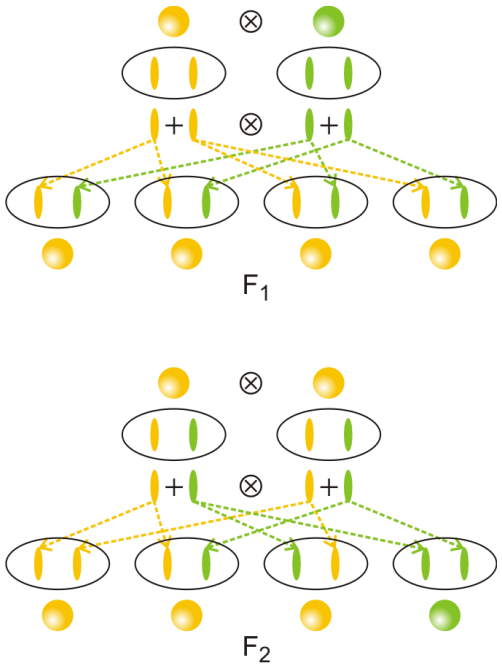


Fig. 1.2 Mendel'l laws of inheritance. The sketch illustrates Mendel's laws of inheritance: (i) the law of segregation and (ii) the law of independent assortment. Every (diploid) organism carries two copies of each gene, which are separated during the process of reproduction. Every offspring receives one randomly chosen copy of the gene from each parent. Encircled are the genotypes formed from two alleles, yellow or green, and above or below the genotypes are the phenotypes expressed as the colors of seeds of the garden pea (*pisum sativum*). The upper part of the figure shows the first generation (F_1) of progeny of two homozygous parents – parents who carry two identical alleles. All genotypes are heterozygous and carry one copy of each allele. The yellow allele is dominant and hence the phenotype expresses yellow color. Crossing two F_1 individuals (lower part of the figure) leads to two homozygous and two heterozygous offspring. Dominance causes the two heterozygous genotypes and one homozygote to develop the dominant phenotype and accordingly the observable ratio of the two phenotypes in the F_2 generation is 3:1 on the average as observed by Gregor Mendel in his statistics of fertilization experiments (see table 1.1).

chance that two celebrate birthday one the same day; 57 yield more than 99% and 70 persons exceed 99,9%. An implicit assumption in this calculation has been that births are uniformly distributed over the year or, in other words, the probability that somebody has the birthday on some day does not depend on the particular day of the year. In mathematical statistics such an assumption is called a *null hypothesis* (see [143] and section 2.6.2).

Laws in classical physics are considered as deterministic in the sense that a single measurement is expected to yield a precise result, deviations from which are interpreted as lack in precision of the machinery used. Random scatter when it is observed is thought to be caused by variation in not sufficiently well controlled experimental conditions. Apart from deterministic laws other regularities are observed in nature, which become evident only when sample sizes are made sufficiently large through repetition of experiments. It is appropriate to call such regularities *statistical laws*. Statistics in biology of plant inheritance has been pioneered by the Augustinian monk Gregor Mendel who discovered regularities in the progeny of the *garden pea* in controlled fertilization experiments [324] (figure 1.2). As a third and final example we consider some of Mendel's data in order to illustrate a statistical law. In table 1.1 the results of two typical experiments distinguishing roundish or wrinkled seeds of yellow or green color are listed. The ratios observed with single plants exhibit large scatter. In the mean values for ten plants shown in the table some averaging has occurred but still the deviations from the ideal values are recognizable. Mendel carefully investigated several hundred plants and then the statistical law of inheritance demanding a ratio of 3:1 became

Table 1.1 Statistics of Gregor Mendel's experiments with the garden pea (*pisum sativum*)**.** In total Mendel analyzed 7324 seeds from 253 hybrid plants in the second trial year, 5474 were round or roundish and 1850 angular wrinkled yielding a ratio 2.96:1. The color was recorded for 8023 seeds from 258 plants out of which 6022 were yellow and 2001 were green with a ratio of 3.01:1. The results of two typical experiments with ten plants, which deviate stronger because of the smaller sample size are shown in the table.

	Form of seed			Color of seeds		
plants	round	wrinkled	ratio	yellow	green	ratio
1	45	12	3.75	25	11	2.27
2	27	8	3.38	32	7	4.57
3	24	7	3.43	14	5	2.80
4	19	10	1.90	70	27	2.59
5	32	11	2.91	24	13	1.85
6	26	6	4.33	20	6	3.33
7	88	24	3.67	32	13	2.46
8	22	10	2.20	44	9	4.89
9	28	6	4.67	50	14	3.57
10	25	7	3.57	44	18	2.44
total	336	101	**3.33**	355	123	**2.89**

evident [324].[8] Ronald Fisher in a somewhat polemic publication [142] reanalyzed Mendel's experiments, questioned Mendel's statistics, and accused him of intentionally manipulating his data because the results are too close to the ideal ratio. Fisher's publication initiated a long lasting debate during which many scientists spoke up in favor of Mendel [348, 349] but there were also critical voices saying that most likely Mendel has unconsciously or consciously eliminated extreme outliers [102]. In 2008 a recent book declared *the end of the Mendel-Fisher controversy* [150]. In section 2.6.2 we shall discuss statistical laws and Mendel's experiments in the light of present day mathematical statistics by applying the so-called χ^2 test.

Probability theory in its classical form is more than three hundred years old. Not accidentally the concept arose in thinking about gambling, which was considered as a domain of chance in contrast to rigorous science. It took indeed rather long time before the concept of probability entered scientific thought in the nineteenth century. The main obstacle for the acceptance of probabilities in physics was the strong belief in determinism that has not been overcome before the advent of quantum theory. Probabilistic concepts in physics of the nineteenth century were still based on deterministic thinking, although the details of individual events were considered to be too numerous to be accessible to calculation at the microscopic level. It is worth mentioning that thinking in terms of probabilities entered biology earlier, already in the second half of the nineteenth century through the reported works on the genetics of inheritance by Gregor Mendel and the considerations about pedigrees by Francis Galton (see section 5.2.2). The reason for this difference appears to lie in the very nature of biology: Small sample sizes are typical, most of the regularities are probabilistic and become observable only through the application of probability theory. Ironically, Mendel's investigations and papers did not attract a broad scientific audience before their *rediscovery* at the beginning of the twentieth century. The scientific community in the second half of the nineteenth century was simply not yet prepared for the acceptance of quantitative and moreover probabilistic concepts in biology.

Classical probability theory is dealing successfully with a number of concepts like conditional probabilities, probability distributions, moments and others, which shall be presented in the next section making use of set theoretic concepts that can provide much deeper insight into the structure of probability theory than mere counting. In addition, the more elaborate notion of probability derived from set theory is absolutely necessary for extrapolation to infinitely large and uncountable sample sizes. Uncountability is the unavoidable attribute of sets derived from continuous variables, and by means of the set theoretic approach real numbers, $\mathbf{x} \in \mathbb{R}^n$, become accessible to probability measures. From now on we shall use only the set theoretic con-

[8] According to modern genetics this ratio as well as other ratios between distinct inherited phenotypes are idealized values that are found only for completely independent genes [180], which lie either on different chromosomes or sufficiently far apart on the same chromosome.

cept, because it can be introduced straightforwardly for countable sets and discrete variables and, in addition, it can be extended to probability measures for continuous variables.

1.3 Interpretations of probability

Before introducing the currently most popular and standard theory of probability we make a brief digression into the dominant philosophical interpretations: (i) the classical interpretations that we adopted in section 1.2, (ii) the frequency-based interpretation that will be in the background of the rest of the book, and (iii) the Bayesian or *subjective* interpretation.

The *classical interpretation of probability* goes back to the concepts and works of the Swiss mathematician Jakob Bernoulli and the French mathematician and physicist Pierre-Simon Laplace, who has been the first presenting a clear definition of probability [272, pp. 6-7]:

> *"The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all cases possible."*

Clearly, this definition is tantamount to equation (1.1) and the explicitly stated assumption of equal probabilities is now called *principle of indifference*. This classical definition of probability has been questioned during the nineteenth century among others by the two English logicians and philosophers George Boole [47] and John Venn [446], who among others initiated a paradigm shift from the classical view to the modern frequency interpretations of probabilities.

The modern interpretations of the concept of probabilities fall essentially into two categories that can be characterized as *physical probabilities* and *evidential probabilities* [186]. Physical probabilities are often called *objective* or frequency-based probabilities and their proponents are addressed as *frequentists*. Influential proponents of the frequency-based probability theory were, besides the pioneer John Venn, the Polish American mathematician Jerzy Neyman, the English statistician Egon Pearson, the English statistician and theoretical biologist Ronald Fisher, the Austro-Hungarian American mathematician and scientist Richard von Mises and the German American philosopher of science Hans Reichenbach. The physical probabilities are derived from some real process like radioactive decay, chemical reaction, turning a roulette wheel, or rolling dice. In all such systems the notion of probability makes sense only when it refers to some well defined experiment with a random component. Frequentism comes in two versions: (i) *finite frequentism* and (ii) *hypothetical frequentism*. Finite frequentism replaces the notion of *'total number of events'* in equation (1.1) by *'actually recorded number of events'* and is thus congenial to philosophers with empiricist scruples. Philosophers have a number of problems with finite frequentism, we mention for example the

small sample problems: One can never speak about the probability of a single experiment and there are cases of unrepeated and unrepeatable experiments. A coin that is tossed exactly once yields a relative frequency of heads of zero or one, no matter what its bias really is. Another famous example is the spontaneous radioactive decay of an atom where the probabilities of decaying follow a continuous exponential law but according to finite frequentism it decays with probability one at its actual decay time. The evolution of the universe or the origin of life can serve as cases of unrepeatable experiments, but people like to speak about the probability that the development has been such or such. Personally, I think it would do no harm to replace *probability* by *plausibility* in such estimates concerned with unrepeatable single events.

Hypothetical frequentism complements the empiricism of finite frequentism by the admission of infinite sequences of trials. Let N be the total number of repetitions of an experiment and n_A the number of trials when the event A has been observed, then the relative frequency of recording the event A is an *approximation of the probability* for the occurrence of A:

$$\text{Prob}\,(A) \;=\; P(A) \;\approx\; \frac{n_A}{N}\;.$$

This equation is essentially the same as (1.1) but the claim of the hypothetical frequentists' interpretation is that there exists a *true frequency* or *true probability* to which the relative frequency converged when we repeated the experiment an infinite number of times[9]

$$P(A) \;=\; \lim_{N\to\infty}\frac{n_A}{N} \;=\; \frac{|A|}{|\Omega|} \quad \text{with} \quad A\in\Omega. \tag{1.2}$$

The probability of an event A relative to a sample space Ω is then defined as the limiting frequency of A in Ω. As N goes to infinity $|\Omega|$ becomes infinitely large and depending on whether $|A|$ is finite or infinite $P(A)$ is either zero or may adopt a nonzero limiting frequency. It is based on two a priori assumptions that have the character of axioms:

(i) *Convergence:* For any event A exists a limiting relative frequency, the probability $P(A)$ that fulfils $0 \le P(A) \le 1$.
(ii) *Randomness:* The limiting relative frequency of each event in a collective Ω is the same in any *typical* infinite subsequence of Ω.

A typical sequence is *sufficiently random*[10] in order to avoid results biased by predetermined order. As a negative example of an acceptable sequence we consider *'head, head, head, head, ...'* recorded by tossing a coin. If it was

[9] The absolute value symbol, '$|A|$', means here the size or cardinality of A being the number of elements in set A (section 1.4).

[10] Sequences are sufficiently random when they are obtained through recordings of random events. *Random sequences* are approximated by the sequential outputs of *pseudorandom number generators*. 'Pseudorandom' implies here that the approximately random sequence is created by some deterministic, i.e. nonrandom, algorithm.

obtained with a fair coin – not with a coin with two heads – $|A|$ is 1 and $P(A) = 1/|\Omega| = 0$, and we may say this particular events is of measure zero and the sequence is not typical. The sequence *'head, tail, head, tail, ...'* is not typical as well despite the fact that it yields the same probabilities as a fair coin. We should be aware that the extension to infinite series of experiments leaves the realm of empiricism and caused thoroughbred philosophers to reject the claim that the interpretation of probabilities by hypothetical frequentism is more *objective* than others.

Nevertheless, frequentist probability theory is not in conflict with the mathematical axiomatization of probability theory and it provides straightforward guidance in applications to real-world problems. The pragmatic view that stands at the beginning of the dominant concept in current probability theory has been phrased nicely by William Feller, the Croatian-American mathematician and author of the classic introduction to probability theory in two volumes [133, 134, Vol.I, pp. 4-5]:

> *"The success of the modern mathematical theory of probability is bought at a price: the theory is limited to one particular aspect of 'chance'. ... we are not concerned with modes of inductive reasoning but with something that might be called physical or statistical probability."*

He also expresses clearly his attitude towards pedantic scruples of philosophic purists:

> *"..., in analyzing the coin tossing game we are not concerned with the accidental circumstances of an actual experiment, the object of our theory is sequences or arrangements of symbols such as 'head, head, tail, head, ...'. There is no place in our system for speculations concerning the probability that the sun will rise tomorrow. Before speaking of it we should have to agree on an idealized model which would presumably run along the lines 'out of infinitely many worlds one is selected at random ...'. Little imagination is required to construct such a model, but it appears both uninteresting and meaningless."*

We shall adopt the frequentist interpretation throughout this monograph but mention here briefly two more interpretations of probability in order to show that it is not the only reasonable probability theory.

The *propensity interpretation* of probability was proposed by the American philosopher Charles Peirce in 1910 [365] and reinvented by Karl Popper [371, pp. 65-70] (see also [372]) more than forty years later [186, 330]. *Propensity* is a tendency to do or to achieve something and in relation to probability, propensity means that it makes sense to talk about the probabilities of single events. As an example we mention the probability – propensity – of a radioactive atom to decay within the next one thousand years, and thereby we make a conclusion from the behavior of an ensemble to a single member of the ensemble. For a fair coin we might say that it has a probability of $\frac{1}{2}$ to score 'head' when tossed, and precisely expressed we should say that the

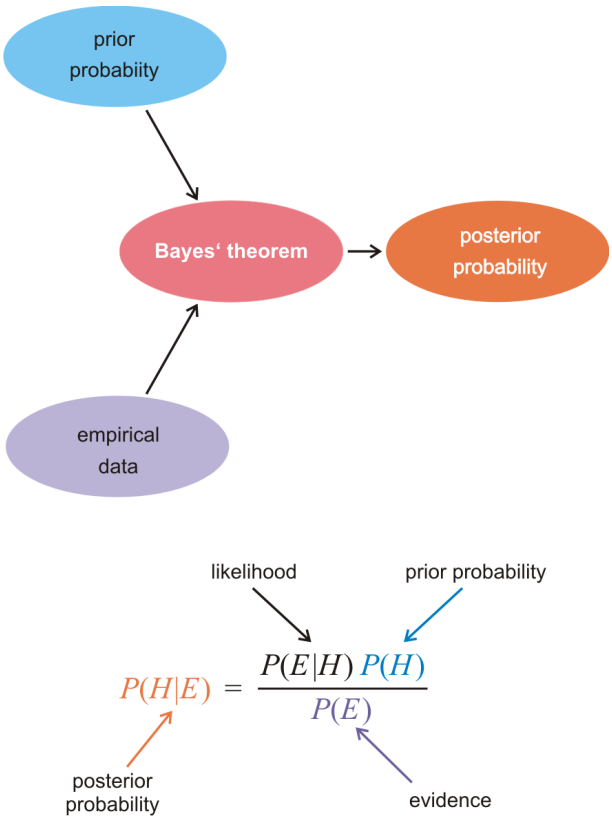


Fig. 1.3 A sketch of the Bayesian method. Prior information of probabilities is confronted with empirical data and converted into a new distribution of probabilities by means of Bayes' theorem according to the formula shown above [98, 415].

coin has the propensity to yield a sequence of outcomes, in which the limiting frequency of scoring 'heads' is $\frac{1}{2}$. The single case propensity is accompanied by, but distinguished from, the *long-run* propensity [175]:

> *"A long-run propensity theory is one in which propensities are associated with repeatable conditions, and are regarded as propensities to produce in a long series of repetitions of these conditions frequencies, which are approximately equal to the probabilities.*

Long-run in these theories is still distinct from infinitely long run in order to avoid basic philosophical problems. As it looks, the use of *propensities* rather than *frequencies* constitutes a language that is somewhat more careful and hence more acceptable in philosophy than the frequentist interpretation.

Finally, we sketch the most popular example of a theory based on *evidential probabilities*: Bayesian statistics, named after the eighteenth century English mathematician and Presbyterian minister Thomas Bayes. In contrast to the frequentists' view probabilities are *subjective* and exist only in the human mind. From a practitioner's point of view one major advantage of the Bayesian approach is the direct insight into the process of improving the knowledge on the object of investigation. In order to understand Bayes' theorem we need the notion of conditional probabilities (for a precise definition see section 1.6.4): For a conditional probability the reference ensemble is not the entire sample space Ω but some event, say B. Then, we have

$$P(A|B) \;=\; \frac{P(A \text{ and } B)}{P(B)} \;=\; \frac{P(AB)}{P(B)}\,, \tag{1.3}$$

where 'A and B' indicates the joint probability of both events A and B.[11] The conditional probability $P(A|B)$ is obtained as the probability of the simultaneous occurrence of events A and B divided by the probability of the occurrence of B alone. If the event B is the entire sample space, $B \equiv \Omega$ we obtain:

$$P(A|\Omega) \;=\; \frac{P(A \text{ and } \Omega)}{P(\Omega)} \;=\; \frac{P(A\Omega)}{P(\Omega)} \;=\; \frac{P(A)}{1} \;=\; P(A)\,,$$

the conditional probability is equal to the unconditioned probability. Conditional probabilities can be inverted straightforwardly in the sense that we ask about the probability of B under the condition that event A has occurred:

$$P(B|A) \;=\; \frac{P(A \text{ and } B)}{P(A)} \;=\; \frac{P(AB)}{P(A)} \quad \text{since} \quad P(AB) \,=\, P(BA)\,, \tag{1.3'}$$

which implies $P(A|B) \neq P(B|A)$ unless $P(A) = P(B)$. In other words the conditional probability can be readily inverted, and as expected $P(A|B)$ and $P(B|A)$ are on equal footing in probability theory. Calculation of $P(AB)$ from both equations, (1.3) and (1.3'), and setting the expressions equal yields

$$P(A|B)\,P(B) \,=\, P(AB) \,=\, P(B|A)\,P(A) \;\Longrightarrow\; P(B|A) \,=\, P(A|B)\,\frac{P(B)}{P(A)}\,,$$

which properly interpreted represents Bayes' theorem.

Bayes'theorem provides a straightforward interpretation of conditional probabilities and their inversion in terms of models or hypothesis (H) and data (E). The conditional probability $P(E|H)$ corresponds to the conventional procedure in science: Given a set of hypothesis cast into a model H the task is to calculate the probabilities of the different outcomes E. In physics

[11] From the next section 1.4 on we shall use the set theoretic symbol intersection, '$\cap$', instead of 'and'; AB is an abbreviated notation for 'A and B'.

and chemistry, where we are dealing with well established theories and models, this is, in essence, the common situation. Biology, economics, social sciences and other disciplines, however, are often confronted with situations where no confirmed models exist and then we want to test and improve the probability of a model. We need to invert the conditional probability since we are interested in testing the model in the light of the data available or, in other words, the conditional probability $P(H|E)$ becomes important: What is the probability that a hypothesis H is justified given a set of measured data encapsulated in evidence E? The Bayesian approach casts equations (1.3) and (1.3') into Bayes' theorem,

$$P(H|E) \;=\; P(E|H)\,\frac{P(H)}{P(E)} \;=\; \frac{P(E|H)}{P(E)}\cdot P(H)\;, \tag{1.4}$$

and provides a hint on how to proceed – at least in principle (figure 1.3). An *prior probability* in form of a hypothesis $P(H)$ is converted into evidence according to the likelihood principle $P(E|H)$. The basis of the prior understood as all *a priori* knowledge comes form many sources: theory, previous experiments, gut feeling, etc. New empirical information is incorporated in the *inverse probability* computation from data to model, $P(H|E)$, yielding thereby the improved *posterior probability*. The advantage of the Bayesian approach is that a change of opinion in the light of new data is *'part of the game'*. In general, parameters are input quantities of frequentist statistics and if unknown assumed to be available through consecutive repetition of experiments, whereas they are understood as random variables in the Bayesian approach. The direct application of the Bayesian theorem in practice involves quite elaborate computations that were not possible in real world examples before the advent of electronic computers. An example of the Bayesian approach and the calculations involved thereby is presented in section 2.6.4.

Bayesian statistics has become popular in disciplines where model building is a major issue. Examples from biology are among others bioinformatics, molecular genetics, modeling of ecosystems, and forensics. Bayesian statistics is described in a large number of monographs, for example, in references [77, 159, 234, 275].

1.4 Sets and sample spaces

Conventional probability theory is based on several axioms that are rooted in set theory, which will be introduced and illustrated in this section. The development of set theory in the eighteen seventieth was initiated by Georg Cantor and Richard Dedekind and provided the possibility to build among many other things the concept of probability on a firm basis that allows for an extension to certain families of uncountable samples as they occur, for example, with continuous variables. Present day probability theory thus can be understood as a convenient extension of the classical concept by means of set and measure theory. We begin by repeating a few indispensable notions and operations of set theory.

Sets are collections of objects with two restrictions: (i) Each object belongs to one set and cannot be a member of two or more sets, and (ii) a member of a set must not appear twice or more often. In other words, objects are assigned to sets unambiguously. In the application to probability theory we shall denote the *elementary objects* by the lower case Greek letter *omega*, ω – if necessary with various sub- and superscripts – and call them *sample points* or *individual results*. The collection of all objects ω under consideration, the *sample space*, is denoted by the upper case Greek letter Ω with $\omega \in \Omega$. *Events*, A, are subsets of sample points that fulfil some condition[12]

$$A \;=\; \big\{\boldsymbol{\omega}, \omega_k \in \Omega :\; f(\boldsymbol{\omega}) = c\big\} \tag{1.5}$$

with $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots)$ being the set of individual results which fulfil the condition $f(\boldsymbol{\omega}) = c$.

Next we repeat the basic logical operations with sets. Any partial collection of points $\omega_k \in \Omega$ is a *subset* of Ω. We shall be dealing with fixed Ω and, for simplicity, often call these subsets of Ω just sets. There are two extreme cases, the entire sample space Ω and the *empty set*, $\emptyset$. The number of points in a set S is called its size or *cardinality* written as $|S|$, and thus $|S|$ is a nonnegative integer or infinity. In particular, the size of the empty set is $|\emptyset| = 0$. The unambiguous assignment of points to sets can be expressed by[13]

$$\omega \in S \quad \text{exclusive or} \quad \omega \notin S \;.$$

Consider two sets A and B. If every point of A belongs to B, then A is contained in B. A is a subset of B and B is a superset of A:

---

[12] The meaning of *condition* will become clearer later on. For the moment it is sufficient to understand a condition as a restriction cast in a function $f(\boldsymbol{\omega})$, which implies that not all subsets of sample points belong to A. Such a condition, for example, is a score '6' in rolling two dice, which comprises the five sample points: $A = \{'1+5','2+4','3+3','4+2','5+1'\}$.

[13] In order to be unambiguously clear we shall write *or* for *and/or* and *exclusive or* for *or* in the strict sense.

$$A \subset B \quad \text{and} \quad B \supset A \; .$$

Two sets are identical if the contain exactly the same points and then we write $A = B$. In other words, $A = B$ iff[14] $A \subset B$ and $B \subset A$.

Some basic operations with sets are illustrated in figure 1.4. We briefly repeat them here:

Complement. The complement of the set A is denoted by A^c and consists of all points not belonging to A:[15]

$$A^c \; = \; \{\omega | \omega \notin A\} \; . \tag{1.6}$$

There are three evident relations which can be verified easily: $(A^c)^c = A$, $\Omega^{\,c} = \emptyset$, and $\emptyset^{\,c} = \Omega$.

Union. The union of the two sets A and B, $A \cup B$, is the set of points, which belong to at least one of the two sets:

$$A \cup B \; = \; \{\omega | \omega \in A \text{ or } \omega \in B\} \; . \tag{1.7}$$

Intersection. The intersection of the two sets A and B, $A \cap B$, is the set of points, which belong to both sets:[16]

$$A \cap B \; = \; AB \; = \; \{\omega | \omega \in A \text{ and } \omega \in B\} \; . \tag{1.8}$$

Unions and intersections can be executed in sequence and are also defined for more than two sets, or even for a countably infinite number of sets:

$$\bigcup_{n=1,\ldots} A_n \; = \; A_1 \cup A_2 \cup \cdots \; = \; \{\omega | \omega \in A_n \text{ for at least one value of } n\} \; ,$$

$$\bigcap_{n=1,\ldots} A_n \; = \; A_1 \cap A_2 \cap \cdots \; = \; \{\omega | \omega \in A_n \text{ for all values of } n\} \; .$$

The proof of these relations is straightforward, because the commutative and the associative laws are fulfilled by both operations, intersection and union:

$$A \cup B \; = \; B \cup A \; , \; A \cap B \; = \; B \cap A \; ;$$

$$(A \cup B) \cup C \; = \; A \cup (B \cup C) \; , \; (A \cap B) \cap C \; = \; A \cap (B \cap C) \; .$$

Difference. The set theoretic difference, $A \setminus B$, is the set of points, which belong to A but not to B:

[14] The word *'iff'* stands for *if and only if*.

[15] Since we are considering only fixed sample sets Ω these points are uniquely defined.

[16] For short $A \cap B$ is often written simply as AB.

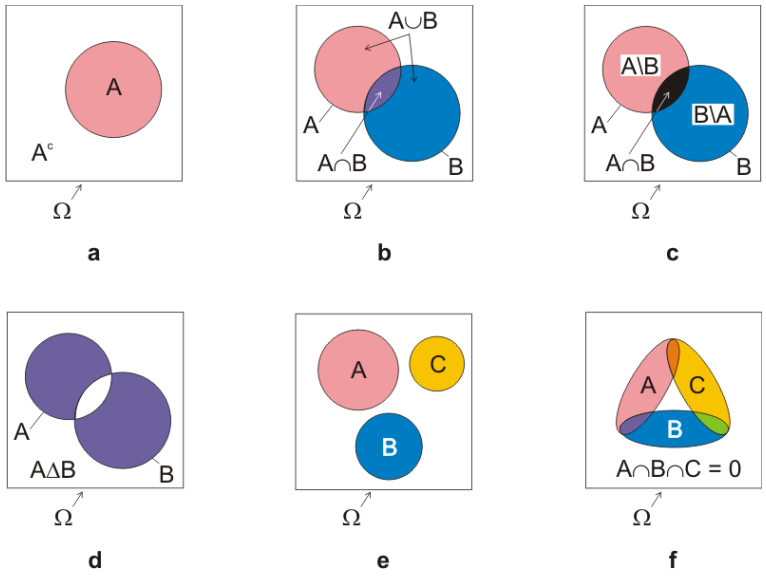


Fig. 1.4 Some definitions and examples from set theory. Part **a** shows the complement A^c of a set A in the sample space Ω. In part **b** we explain the two basic operations union and intersection, $A \cup B$ and $A \cap B$, respectively. Parts **c** and **d** show the set-theoretic difference, $A \setminus B$ and $B \setminus A$, and the symmetric difference, $A \triangle B$. In parts **e** and **f** we demonstrate that a vanishing intersection of three sets does not imply pairwise disjoint sets. The illustrations are made by means of *Venn diagrams* [181, 182, 444, 445].

$$A \setminus B \;=\; A \cap B^c \;=\; \{\omega | \omega \in A \text{ and } \omega \notin B\} \; . \tag{1.9}$$

In case $A \supset B$ we write $A - B$ for $A \setminus B$ and have $A \setminus B = A - (A \cap B)$ as well as $A^c = \Omega - A$.

Symmetric difference. The symmetric difference $A \Delta B$ is the set of points which belongs exactly to one of the two sets A and B. It is used in advanced theory of sets and is symmetric as it fulfils the commutative law, $A \Delta B = B \Delta A$:

$$A \Delta B \;=\; (A \cap B^c) \cup (A^c \cap B) \;=\; (A \setminus B) \cup (B \setminus A) \; . \tag{1.10}$$

Disjoint sets. Disjoint sets A and B have no points in common and hence their intersection, $A \cap B$, is empty. They fulfill the following relations:

$$A \cap B \;=\; \emptyset \; , \quad A \subset B^c \text{ and } B \subset A^c \; . \tag{1.11}$$

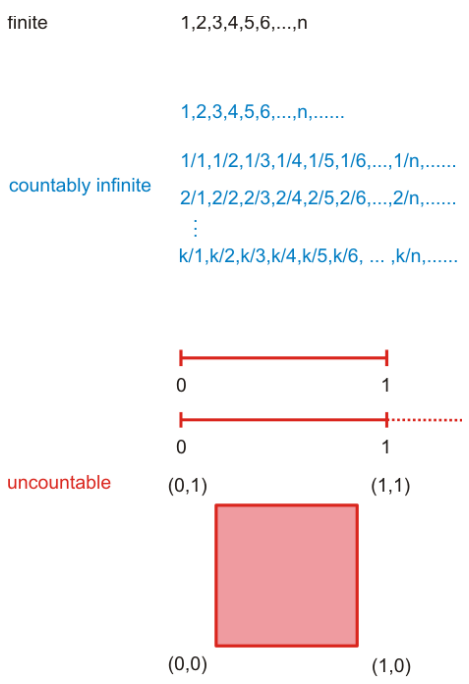


Fig. 1.5 Sizes of sample sets and countability. Finite (black), countably infinite (blue), and uncountable sets (red) are distinguished. We show examples of every class. A set is countably infinite when its elements can be assigned uniquely to the natural numbers ($\mathbb{N}_{>0} = 1,2,3,\ldots,n,\ldots$).

Several sets are disjoint only if they are pairwise disjoint. For three sets, A, B and C, this requires $A \cap B = \emptyset$, $B \cap C = \emptyset$, and $C \cap A = \emptyset$. When two sets are disjoint the addition symbol is (sometimes) used for the union, $A + B$ for $A \cup B$. Clearly we have always the valid *decomposition*: $\Omega = A + A^c$.

Sample spaces may contain finite or infinite numbers of sample points. As shown in figure 1.5 it is important to distinguish further between different classes of infinity:[17] *countable* and *uncountable* numbers of points. The set of *rational numbers* $\mathbb{Q}$, for example, is a countably infinite since the numbers can be labeled and assigned uniquely to the positive integers also called *natural numbers* $\mathbb{N}_{>0} : 1 < 2 < 3 < \cdots < n < \cdots$. The set of *real numbers* $\mathbb{R}$, cannot be ordered in such a way and hence it is uncountable (For the notations used for number systems see the appendix 'Notations').

[17] Georg Cantor attributed the cardinality $\aleph_0$ to countably infinite sets and characterized uncountable sets by the sizes $\aleph_1$, $\aleph_2$, etc.

1.5 Probability measure on countable sample spaces

For countable sets it is straightforward and almost trivial to measure the size of the set by counting the numbers of sample points they contain. The ratio

$$P(A) \;=\; \frac{|A|}{|\Omega|} \tag{1.12}$$

gives the probability for the occurrence of event A and the expression is of course identical with equation (1.1) defining the classical probability. For another event, for example B, holds $P(B) = |B|/|\Omega|$. A calculation of the the sum of the two probabilities, $P(A)+P(B)$, requires some care, since figure 1.4 suggests that only an inequality holds (see previous section 1.4):

$$|A| \;+\; |B| \;\geq\; |A \cup B| \;.$$

The excess of $|A|+|B|$ over the size of the union $|A \cup B|$ is precisely the size of the intersection $|A \cap B|$ and thus we find

$$|A| \;+\; |B| \;=\; |A \cup B| \;+\; |A \cap B|$$

or by division through the size of sample space Ω we obtain

$$\begin{aligned} P(A) \,+\, P(B) \;&=\; P(A \cup B) \,+\, P(A \cap B) \quad \text{or} \\ P(A \cup B) \;&=\; P(A) \,+\, P(B) \,-\, P(A \cap B) \;. \end{aligned} \tag{1.13}$$

Only in case the intersection is empty, $A \cap B = \emptyset$, the two sets are disjoint and their probabilities are additive, $|A \cup B| = |A| + |B|$, and hence

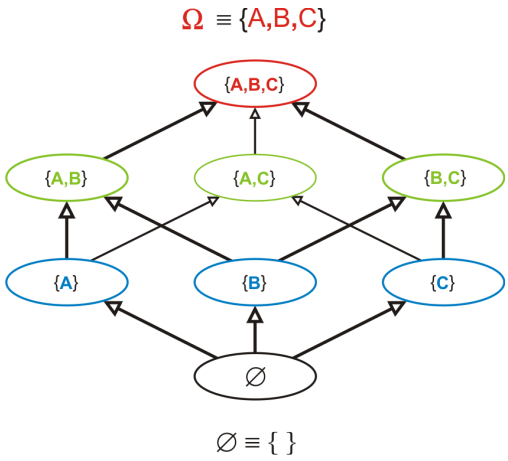


Fig. 1.6 The powerset. The powerset $\Pi(\Omega)$ is a set containing all subsets of Ω including the empty set $\emptyset$ (black) and Ω itself (red). The figure sketches the construction of the powerset for a sample space of three events A, B, and C (single events in blue, and double events in green). The relation between sets and sample points is also illustrated in a set level diagram (see the black and red level in figure 1.15)

$$P(A+B) \;=\; P(A) \,+\, P(B) \;\text{ iff }\; A \cap B = \emptyset \,. \tag{1.14}$$

It is important to memorize this condition for later use, because it represents an implicitly made assumption for computing probabilities.

1.5.1 Probability measure

Now are now in the position to define a probability measure by means of basic axioms of probability theory and we present the three axioms as they were first formulated by Andrey Kolmogorov [259]:

A *probability measure* on the sample space Ω is a function of subsets of Ω, $P : S \to P(S)$, which is defined by the three axioms:

(i) For every set $A \subset \Omega$, the value of the probability measure is a nonnegative number, $P(A) \geq 0$ for all A,
(ii) the probability measure of the entire sample set – as a subset – is equal to one, $P(\Omega) = 1$, and
(iii) for any two disjoint subsets A and B, the value of the probability measure for the union, $A \cup B = A + B$, is equal to the sum of its values for A and B,

$$P(A \cup B) \;=\; P(A+B) \;=\; P(A) + P(B) \text{ provided } P(A \cap B) \;=\; \emptyset \,.$$

Condition (iii) implies that for any countable – eventually infinite – collection of disjoint or non-overlapping sets, A_i, $i = 1, 2, 3, \ldots$, with $A_i \cap A_j = \emptyset$ for all $i \neq j$, a relation called *σ-additivity*

$$P\left(\bigcup_i A_i\right) \;=\; \sum_i P(A_i) \;\text{ or }\; P\left(\sum_{i=1}^{\infty} A_i\right) \;=\; \sum_{i=1}^{\infty} P(A_i) \tag{1.15}$$

is fulfilled.

In other words, the probabilities associated with disjoint sets are additive. Clearly we have also $P(A^c) = 1 - P(A)$, $P(A) = 1 - P(A^c) \leq 1$, and $P(\emptyset) = 0$. For any two sets $A \subset B$ we find $P(A) \leq P(B)$ and $P(B - A) = P(B) - P(A)$, and for any two arbitrary sets A and B we can write the union as a sum of two disjoint sets

$$\begin{aligned} A \cup B \;&=\; A \,+\, A^c \cap B \text{ and} \\ P(A \cup B) \;&=\; P(A) \,+\, P(A^c \cap B) \,. \end{aligned}$$

Since $B \subset A^c \cap B$ we obtain $P(A \cup B) \leq P(A) + P(B)$.

The set of all subsets of Ω is called the *powerset* $\Pi(\Omega)$ (figure 1.6). It contains the empty set $\emptyset$, the entire sample space Ω and the subsets of

Ω, and this includes the results of all set theoretic operations that were listed in the previous section 1.4. We illustrate the relation between the sample point ω, an event A, the sample space Ω and the powerset $\Pi(\Omega)$ by means of an example, the repeated coin toss, which we shall analyze as Bernoulli process in section 3.1.3. Flipping a coin has two outcomes: '$\mathsf{0}$' for *head* and '$\mathsf{1}$' *tail* and one particular coin toss experiment might give the sequence $(\mathsf{0},\mathsf{1},\mathsf{1},\mathsf{1},\mathsf{0},\ldots,\mathsf{1},\mathsf{0},\mathsf{0})$. Thus the sample points ω for flipping the coin n-times are binary n-tuples or strings, $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ with $\omega_i \in \Sigma = \{\mathsf{0},\mathsf{1}\}$.[18] Then, the sample space Ω is the space of all binary strings of length n commonly denoted by Σ^n and it has the cardinality $|\Sigma^n| = 2^n$. The extension to the set of all strings of any finite length is straightforward,

$$\Sigma^* \;=\; \bigcup_{i\in\mathbb{N}} \Sigma^i \;=\; \{\varepsilon\} \cup \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \ldots \;, \tag{1.16}$$

and this set is called *Kleene star* after the American mathematician Stephen Kleene. Herein $\Sigma^0 = \{\varepsilon\}$ with ε denoting the unique string over Σ^0 called the *empty string*, $\Sigma^1 = \{\mathsf{0},\mathsf{1}\}$, $\Sigma^2 = \{\mathsf{00},\mathsf{01},\mathsf{10},\mathsf{11}\}$, etc. The importance of Kleene star is the *closure* property under concatenation of the sets Σ^i[19]

$$\Sigma^m \Sigma^n \;=\; \Sigma^{m+n} \;=\; \{wv | w \in \Sigma^m \,\text{and}\, v \in \Sigma^n\} \;\text{ with }\; m,n > 0\;. \tag{1.17}$$

Concatenation of strings is the operation

$$w = (\mathsf{0001})\,,\; v = (\mathsf{101}) \implies wv = (\mathsf{0001101})\;,$$

which can be extended to concatenation of sets in the sense of equation 1.17:

$$\begin{aligned}\Sigma^1\Sigma^2 &= \{\mathsf{0},\mathsf{1}\}\{\mathsf{00},\mathsf{01},\mathsf{10},\mathsf{11}\} = \\ &= \{\mathsf{000},\mathsf{001},\mathsf{010},\mathsf{011},\mathsf{100},\mathsf{101},\mathsf{110},\mathsf{111}\} = \Sigma^3\end{aligned}$$

The set Kleene star Σ^* is the smallest superset of Σ, which contains the empty string ε and which is closed under the string concatenation operation. Although all individual strings in Σ^* have finite length, the set Σ^* itself, however, is countably infinite. We end this brief excursion into strings and string operations by considering infinite numbers of repeats directly in the sense of Σ^n the space of strings of lengths n, $\omega = (\omega_1, \omega_2, \ldots) = (\omega_i)_{i\in\mathbb{N}}$ with $\omega_i \in \{0,1\}$ in the limit $\lim n \to \infty$, as they are used in the theory of computing. Then $\Omega = \{0,1\}^{\mathbb{N}}$ is the sample space of all infinitely long

[18] There is a trivial but important distinction between strings and sets: In a string the position of an element matters, whereas in a set it does not. The following three sets are identical: $\{1,2,3\} = \{3,1,2\} = \{1,2,2,3\}$. In order to avoid ambiguities strings are written in (normal) parentheses and sets in curly brackets.

[19] Closure under a given operation is an important property of a set that we shall need later on. The natural numbers $\mathbb{N}$, for example, are closed under addition and the integers $\mathbb{Z}$ are closed under addition and substraction.

binary strings, whose countability as can be easily verified: Every binary string represents the binary encoding N_k of a natural number including '0', $N_k \in \mathbb{N}$, and hence Ω is countable as the natural numbers are.

A subset of Ω will be called an *event* A when a probability measure derived from axioms (i), (ii), and (iii) has been assigned. Often, one is not interested in the full detail of a probabilistic result and events can be easily adapted to lumping together sample points. An illustrative example from statistical physics are the microstates in the partition function, which are lumped together according to some macroscopic properties. Here, we ask, for example, for the probability A that n coin flips show *tail* at least s-times or, in other words, yield a score $k \geq s$:

$$A \;=\; \left\{\omega = (\omega_1, \omega_2, \ldots, \omega_n) \in \Omega : \; \sum\nolimits_{i=1}^{n} \omega_i = k \geq s\right\} \;,$$

where the sample space is $\Omega = \{0, 1\}^n$. The task is now to find a system of events Ξ that allows for a consistent assignment of a probability $P(A)$ to all possible events A. For countable sample spaces Ω the powerset $\Pi(\Omega)$ represents such a system Ξ: We characterize $P(A)$ as a probability measure on $\big(\Omega, \Pi(\Omega)\big)$, and the further handling of probabilities following the procedure outlined below is straightforward. In case of uncountable sample spaces Ω the powerset $\Pi(\Omega)$ will turn out to be too large and a more sophisticated procedure is required (section 1.6.4).

1.5.2 Probability weights

So far we have constructed, compared, and analyzed sets but have not yet introduced weights or numbers for application to real world situations. In order to construct a probability measure that can be adapted to calculations on countable sample space, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$, we have to assign a weight ϱ_n to every sample point ω_n that fulfils the conditions

$$\forall\; n: \quad \varrho_n \;\geq\; 0 \;\text{ and }\; \sum_n \varrho_n = 1 \;. \tag{1.18}$$

Then for $P\left(\{\omega_n\}\right) = \varrho_n \;\forall\; n$ the following two equations

$$\begin{aligned} P(A) \;&=\; \sum_{\omega \in A} \varrho(\omega) \;\text{ for }\; A \in \Pi(\Omega) \quad \text{and} \\ \varrho(\omega) \;&=\; P\left(\{\omega\}\right) \;\text{ for }\; \omega \in \Omega \end{aligned} \tag{1.19}$$

represent a bijective relation between the probability measure P on $\big(\Omega, \Pi(\Omega)\big)$ and the sequences $\varrho = \big(\varrho(\omega)\big)_{\omega \in \Omega}$ in [0,1] with $\sum_{\omega \in \Omega} \varrho(\omega) = 1$. Such a sequence is called a discrete probability density.

The function $\varrho(\omega_n) = \varrho_n$ has to be prescribed by some null hypothesis, estimated or determined empirically, because it is the result of factors lying outside mathematics or probability theory. The uniform distribution is commonly adopted as null hypothesis in gambling as well as for many other purposes: The discrete *uniform distribution*, $\mathcal{U}_\Omega$, assumes that all elementary results $\omega \in \Omega$ appear with equal probability and hence $\varrho(\omega) = 1/|\Omega|$.[20] What is meant here by 'elementary' will become clear in the discussion of applications. Throwing more than one die at a time, for example, can be reduced to throwing one die more often.

In science, particularly in physics, chemistry or biology, the correct assignment of probabilities has to meet the conditions of the experimental setup. An simple example from *scientific gambling* will make this point clear: The fact whether a die is fair and shows all its six faces with equal probability, whether it is imperfect, or whether it has been manipulated and shows, for example, the 'six' more frequently then the other faces is a matter of physics and not mathematics. Empirical information – for example, a calibration curve of the faces is determined by doing and recording a few thousand die rolling experiments – replaces the principle of indifference and assumptions like the null hypothesis of a uniform distribution become obsolete.

Although the application of a probability measure in the discrete case is rather straightforward, we illustrate by means of a simple example. With the assumption of the uniform distribution $\mathcal{U}_\Omega$ we can measure the size of sets by counting sample points as illustrated by considering the scores from throws of dice. For one die the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$ and for the *fair* die we make the assumption

$$P(\{k\}) \;=\; \frac{1}{6}\,;\; k = 1, 2, 3, 4, 5, 6\ .$$

that all six outcomes corresponding to the different faces of the die are equally likely. Based on the assumption of $\mathcal{U}_\Omega$ we obtain the probabilities for the outcome of two simultaneously rolled fair dice (figure 1.7). There are $6^2 = 36$ possible outcomes with scores in the range $k = 2, 3, \ldots, 12$ and the most likely outcome is a count of $k = 7$ points because it has the highest multiplicity: $\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$. The probability distribution is shown here as a histogram, an illustration introduced into statistics by Karl Pearson [362]. It has the shape of discretized tent function and is equivalent to the *probability mass function* (pmf) shown in figure 1.11. A generalization to rolling n dice simultaneously is presented in section 1.9.1 and figure 1.22.

---

[20] The assignment of equal probabilities $\frac{1}{n}$ to n mutually exclusive and collectively exhaustive events, which are indistinguishable except for their tags, is known as *principle of insufficient reason* or *principle of indifference* as it was called by the British economist John Maynard Keynes [247, chap.IV, pp.44-70]. In Bayesian probability theory the *a priori* assignment of equal probabilities is characterized as the simplest *non-informative prior* (see section 1.3).

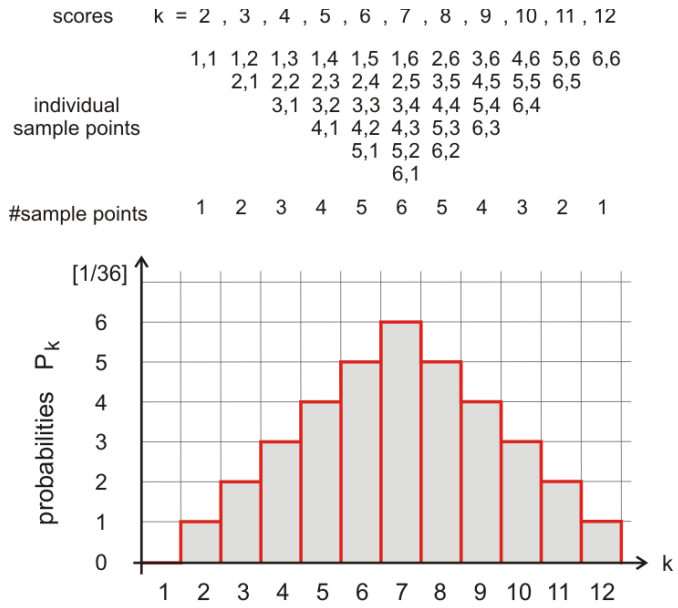


Fig. 1.7 Histogram of the probabilities of throwing two dice. The probability of obtaining two to twelve counts through throwing two perfect or fair dice are based on the equal probability assumption for obtaining the individual faces of a single die. The probability $P(N)$ raises linearly from two to seven and then decreases linearly between seven and twelve: $P(N)$ is a discretized tent map with the additivity or normalization condition $\sum_{k=2}^{12} P(N = k) = 1$. The histogram is equivalent to the probability mass function (pmf) of a random variable $\mathcal{Z}$: $f_{\mathcal{Z}}(x)$ as shown in figure 1.11.

1.6 Discrete random variables and distributions

Conventional deterministic variables are not suitable for descriptions of processes with limited reproducibility. In probability theory and statistics we shall make use of *random* or *stochastic variables*, $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \ldots$, which were invented especially for dealing with random scatter and fluctuations. Even if an experiment is repeated under precisely the same conditions the random variable will commonly take on a different value. The probabilistic nature of random variables is expressed by an equation, which is particularly useful for the definition of probability distribution functions:[21]

$$P_k \;=\; P\big(\mathcal{Z} \,=\, k\big) \quad \text{with} \quad k \,\in\, \mathbb{N}\,. \tag{1.20}$$

[21] Whenever possible we shall use 'k, l, m, n' for discrete counts, $k \in \mathbb{N}$, and 't, x, y, z' for continuous variables, $x \in \mathbb{R}^1$ (see appendix 'Notation').

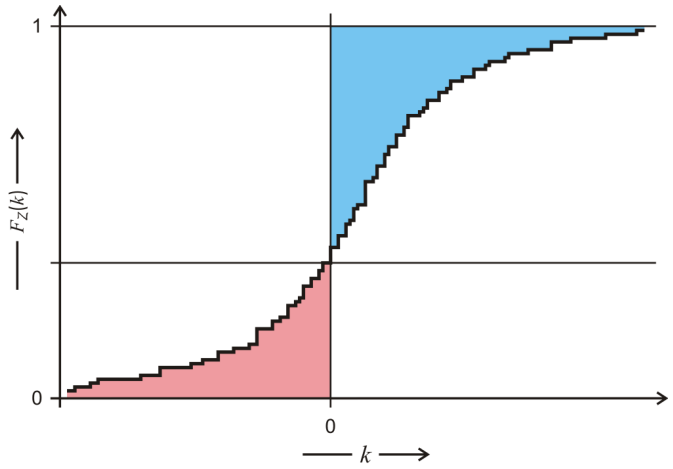


Fig. 1.8 Construction for the calculation of expectation values from cumulative distribution functions. The expectation value calculated from the cumulative distribution function of a discrete variable is obtained as the difference of two contributions: $\sum_{k=0}^{\infty}\big(1 - F_{\mathcal{Z}}(k)\big)$ (blue) and $\sum_{k=-\infty}^{0} F_{\mathcal{Z}}(k)$ (red).

A deterministic variable, $z(t)$, is defined by a function that returns a unique value for a given argument $z(t) = z_t$.[22] In case of the random variable, $\mathcal{Z}(t)$, the single value of the conventional variable has to be replaced by a series of probabilities $P_k(t)$. This series could be visualized, for example, by means of an L_1 normalized probability vector with the probabilities P_k as components: $\mathbf{P} = \big(P_0, P_1, \ldots\big)$ with $\|\mathbf{P}\|_1 = \sum_k P_k = 1$.[23]

1.6.1 Distributions and expectation values

In probability theory the characterization of a random variable is made by a probability distribution function rather than by a vector, because these functions can be applied with minor modifications to the discrete and the continuous case. Two probability functions are particularly important and in general use (see section 1.6.3): the probability mass function (pmf; see figure 1.11)

[22] We use here t as independent variable of the function but do not necessarily imply that t is always time.

[23] The notation of vectors and matrices as used in this text is described in the appendix 'Notation'.

$$f_{\mathcal{Z}}(x) \;=\; \begin{cases} P(\mathcal{Z} = k) \;=\; P_k & \forall\;\; x = k \in \mathbb{N}\,, \\ 0 & \text{anywhere else}\,. \end{cases}$$

or by the *cumulative distribution function* (cdf; see figure 1.12)

$$F_{\mathcal{Z}}(x) \;=\; P(\mathcal{Z} \leq k) \;=\; P(\mathcal{Z} \leq k) \;=\; \sum_{i \leq k} P_i\;.$$

The probability mass function $f_{\mathcal{Z}}(x)$ is not a function in the usual sense, because it has the value zero almost everywhere except at the points $x = k \in \mathbb{N}$ and in this aspect it is closely related to the Dirac delta function (section 1.6.3). Two properties of the cumulative distribution function follow directly from the property of probabilities:

$$\lim_{k \to -\infty} F_{\mathcal{Z}}(k) \;=\; 0 \quad \text{and} \quad \lim_{k \to +\infty} F_{\mathcal{Z}}(k) \;=\; 1\;.$$

The limit at low k-values is chosen in analogy to definitions used later on: Taking $-\infty$ instead of zero as lower limit makes no difference, because $f_{\mathcal{Z}}(-|k|) = P_{-|k|} = 0$ ($k \in \mathbb{N}$) or negative particle numbers have zero probability. Simple examples of probability functions are shown in figures 1.11 and 1.12.

All measurable quantities, for example expectation values and variances, can be computed equally well from either one of the probability functions

$$\mathrm{E}(\mathcal{Z}) \;=\; \sum_{k=-\infty}^{+\infty} k\, f_{\mathcal{Z}}(k) \;=\; \sum_{k=0}^{+\infty} \big(1 - F_{\mathcal{Z}}(k)\big) \;-\; \sum_{k=-\infty}^{0} F_{\mathcal{Z}}(k)\;, \tag{1.21a}$$

$$\begin{aligned} \mathrm{var}(\mathcal{Z}) \;&=\; \sum_{k=-\infty}^{k=+\infty} k^2\, f_{\mathcal{Z}}(k) \;-\; \mathrm{E}(\mathcal{Z})^2 \;= \\ &=\; 2 \sum_{k=0}^{+\infty} k \,\big(1 - F_{\mathcal{Z}}(k)\big) \;-\; \mathrm{E}(\mathcal{Z})^2\;. \end{aligned} \tag{1.21b}$$

In both equations the expressions calculating directly from the cumulative distribution function are valid only for exclusively nonnegative random variables, $\mathcal{Z} \in \mathbb{N}$. The extension to the full domain of discrete random variables, $\mathcal{Z} \in \mathbb{Z}$ is shown below and visualized in figure 1.8.

As an example for the usage of the cumulative distribution function we present a proof for the computation of the expectation values for positive random variables: $\mathrm{E}(\mathcal{Z}) = \sum_{k=0}^{\infty} \big(1 - F_{\mathcal{Z}}(k)\big)$.[24] We show the validity of the expression $\mathrm{E}(\mathcal{Z}) = \sum_{k=1}^{\infty} P(\mathcal{Z} \geq k)$ with $k \in \mathbb{N}$ expanding the '$\geq$' relation and interchanging the order of summations,

[24] The proof is taken from `http://en.wikipedia.org/wiki/Expected_value`, retrieved March 16, 2014.

$$\sum_{k=1}^{\infty} P(\mathcal{Z} \geq k) \;=\; \sum_{k=1}^{\infty}\sum_{j=k}^{\infty} P(\mathcal{Z}=j) \;=\; \sum_{j=1}^{\infty}\sum_{k=1}^{j} P(\mathcal{Z}=j) \;=$$

$$= \sum_{j=1}^{\infty}\sum_{k=1}^{j} P_j \;=\; \sum_{j=1}^{\infty} j\, P_j \;=\; \mathrm{E}(\mathcal{Z})\;,$$

and we introduce now the cumulative distribution function:

$$F_{\mathcal{Z}}(k) \;=\; P(\mathcal{Z} \leq k) \;=\; 1 - P(\mathcal{Z} > k)\;,$$

$$F_{\mathcal{Z}}(k-1) \;=\; P(\mathcal{Z} \leq k-1) \;=\; 1 - P(\mathcal{Z} > k-1) \;=\; 1 - P(\mathcal{Z} \geq k) \;\text{ and}$$

$$\mathrm{E}(\mathcal{Z}) \;=\; \sum_{k=0}^{\infty}\big(1 \;-\; F_{\mathcal{Z}}(k)\big)\;. \qquad \square$$

The generalization of the expectation value to the the negative range yields

$$\mathrm{E}(\mathcal{Z}) \;=\; \sum_{k=0}^{+\infty}\big(1 - F_{\mathcal{Z}}(k)\big) \;-\; \sum_{k=-\infty}^{0} F_{\mathcal{Z}}(k)\;. \tag{1.21c}$$

The partitioning of $\mathrm{E}(\mathcal{Z})$ into a positive and a negative part is visualized in figure 1.8, the derivation of the expression will be given for the continuous case (section 1.9.1).

1.6.2 Random variables and continuity

Random variables on countable sample spaces require a *probability triple* $(\Omega, \Pi(\Omega), P)$ for a precise definition: Ω contains the sample points or individual results, the powerset $\Pi(\Omega)$ provides the events A as subsets, and P finally represents a probability measure as it has been introduced in equation (1.19). Based on such a probability triple we can now define the *random variable* as a numerically valued function $\mathcal{Z}$ of ω on the domain of the entire sample space Ω,

$$\omega \in \Omega\;:\;\; \omega \;\to\; \mathcal{Z}(\omega)\;. \tag{1.22}$$

Random variables, $\mathcal{X}(\omega)$ and $\mathcal{Y}(\omega)$, can be manipulated by conventional operations to yield other random variables, such as

$$\mathcal{X}(\omega) + \mathcal{Y}(\omega)\,,\; \mathcal{X}(\omega) - \mathcal{Y}(\omega)\,,\; \mathcal{X}(\omega)\mathcal{Y}(\omega)\,,\; \mathcal{X}(\omega)/\mathcal{Y}(\omega)\;\; [\mathcal{Y}(\omega) \neq 0]\,,$$

and, in particular, any linear combination of random variables such as $\alpha\mathcal{X}(\omega) + \beta\mathcal{Y}(\omega)$ is a random variable too. Similarly as a function of a function is still a function, and hence a function of a random variable is a random variable,

$$\omega \in \Omega:\;\; \omega \;\to\; \varphi\,(\mathcal{X}(\omega), \mathcal{Y}(\omega)) \;=\; \varphi(\mathcal{X}, \mathcal{Y})\;.$$

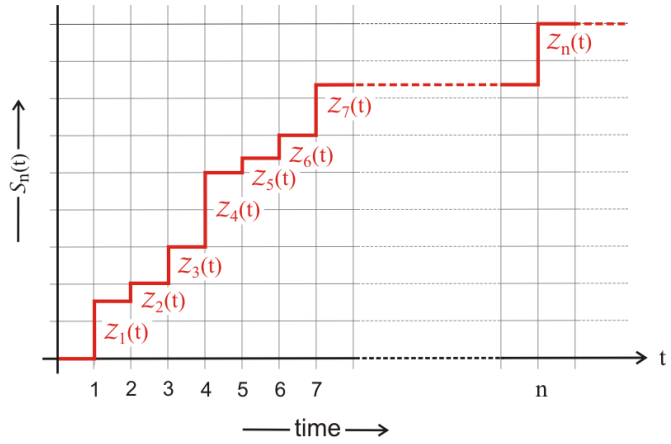


Fig. 1.9 Ordered partial sum of random variables. The sum of random variables, $\mathcal{S}_n(t) = \sum_{k=1}^{n} \mathcal{Z}_k(t)$, represents the cumulative outcome of a series of events described by a class of random variables, $\mathcal{Z}_k$. The series can be extended to $+\infty$ and such cases will be encountered, for example, with probability distributions. The ordering criterion specified in this sketch is *time* t, and we are dealing with a stochastic process, here a jump process. The time intervals need not be constant as shown here. The ordering criterion could be equally well a *spatial coordinate* x, y or z.

Particularly important cases of derived quantities are the partial sums of variables:[25]

$$\mathcal{S}_n(\omega) \;=\; \mathcal{Z}_1(\omega) + \ldots + \mathcal{Z}_n(\omega) \;=\; \sum_{k=1}^{n} \mathcal{Z}_k(\omega)\;. \tag{1.23}$$

Such a partial sum $\mathcal{S}_n$ could be, for example, the cumulative outcome of n successive throws of a die. The series, in principle, could be extended to infinity covering thereby the entire sample space, and then the conservation relation of probabilities, $\mathcal{S}_n = \sum_{k=1}^{\infty} \mathcal{Z}_k = 1$, has to be fulfilled. The terms in the sum can be permuted arbitrarily since no ordering criterium has been introduced so far. Most frequently and in particular in the context of stochastic processes, events will be ordered according to their time t of occurrence (see chapter 3). An ordered series of events where the current cumulative outcome is given by the sum $\mathcal{S}_n(t) = \sum_{k=1}^{n} \mathcal{Z}_k(t)$ is shown in figure 1.9: The plot of the random variable, $\mathcal{S}(t)$, is a multi-step function over a continuous time axis t.

[25] The use of *partial* in this context expresses the fact that the sum need not cover the entire sample space at least not for the moment. Series of rolling dice, for example, could be continued in the future.

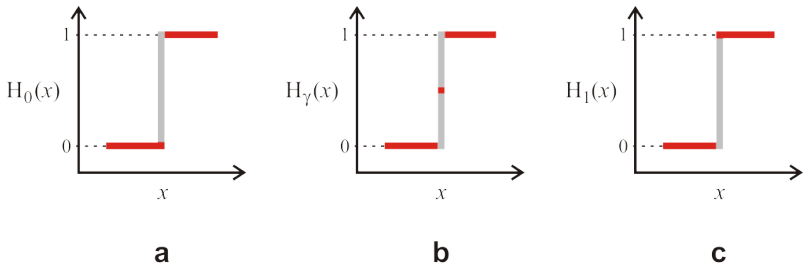


Fig. 1.10 Continuity in probability theory and step processes. Three possible choices of partial continuity or no continuity are shown for the step of the Heaviside function $H_\gamma(x)$: (i) $\gamma = 0$ with left-hand continuity (**a**), (ii) $\gamma \notin \{0, 1\}$ implying no continuity (**b**), and (iii) $\gamma = 1$ with right-hand continuity (**c**). The step function in (**a**) is left-hand semi-differentiable, the step function in (**c**) is right-hand semi-differentiable, and the step function in (**b**) is neither right-hand nor left-hand semi-differentiable. Choice (ii) with $\gamma = 1/2$ allows for making use of the inherent symmetry of the Heaviside function. Choice (iii) is the standard assumption in Lebesgue-Stieltjes integration, probability theory, and stochastic processes. It is also known as *càdlàg*-property (section 3.1.3.3).

Continuity. Steps are inherent discontinuities, and without some further convention we do not know how the value at the step is handled by various step functions. In order to avoid ambiguities, which concern not only the value of the function but also the problem of partial continuity or discontinuity, a convention prior to usage is needed that makes expressions like (1.22) or (1.23) precise. The Heaviside step or Θ function is defined by:

$$\mathrm{H}(x) \;=\; \begin{cases} 0\,, & \text{if } x < 0\,, \\ \text{undefined}\,, & \text{if } x = 0\,, \\ 1\,, & \text{if } x > 0\,. \end{cases} \tag{1.24}$$

It has the discontinuity at the origin $x = 0$ and is undefined there. The Heaviside step function can be interpreted as the integral of the Dirac delta function

$$\mathrm{H}(x) \;=\; \int_{-\infty}^{x} \delta(\xi)\,\mathrm{d}\xi\ ,$$

and this expression becomes ambiguous or meaningless for $x = 0$ as well. The ambiguity can be removed by specifying the value at the origin

$$\mathrm{H}_\gamma(x) \;=\; \begin{cases} 0\,, & \text{if } x < 0\,, \\ \gamma \in [0, 1] & \text{if } x = 0\,, \\ 1\,, & \text{if } x > 0\,. \end{cases} \tag{1.25}$$

In particular, the three definitions shown in figure 1.10 for the value of the function at the step are common.

For a general step function $F(x)$ with the step at x_0 – discrete cumulative probability distributions $F_{\mathcal{Z}}(x)$ may serve as examples – the three possible definitions of the discontinuity at x_0 are expressed in terms of the values (immediately) below and immediately above the step, which we denote by f_{low} and f_{high}, respectively:

(i) $\lim_{\epsilon\to 0} F(x_0-\epsilon) = f_{\text{low}}$ and $\lim_{\epsilon\to\delta>0} F(x_0+\epsilon) = f_{\text{high}}$ with $\epsilon > \delta$ and δ arbitrarily small. The value f_{low} at $x = x_0$ for the function $F(x)$ implies left-hand continuity, the function is semi-differentiable to the left, that is towards decreasing values of x.

(ii) $\lim_{\epsilon\to\delta>0} F(x_0-\epsilon) = f_{\text{low}}$ and $\lim_{\epsilon\to\delta>0} F(x_0+\epsilon) = f_{\text{high}}$ with $\epsilon > \delta$ and δ arbitrarily small, and the value of the step function at $x = x_0$ is neither f_{low} nor f_{high}. Accordingly, $F(x)$ is not differentiable at $x = x_0$. A special definition is chosen in case the inherent inversion symmetry of a step functions should be emphasized: $F(x_0) = \big(f_{\text{low}} + f_{\text{high}}\big)/2$ (see the sign function below).

(iii) $\lim_{\epsilon\to\delta>0} F(x_0-\epsilon) = f_{\text{low}}$ with $\epsilon > \delta$ and δ arbitrarily small and $\lim_{\epsilon\to 0} F(x_0+\epsilon) = f_{\text{high}}$. The value $F(x_0) = f_{\text{high}}$ results in right-hand continuity and semi-differentiability to the right as expressed by *càdlàg*, which is an acronym from French for *'continue à droite, limites à gauche'*. Right-hand continuity is the standard assumption in the theory of stochastic processes. The cumulative distribution functions $F_{\mathcal{Z}}(x)$, for example, are semi-differentiable to the right, that is towards increasing values of x.

A frequently used example of case (ii) is the sign function or signum function, $\operatorname{sgn}(x) = 2\,\mathrm{H}_{\frac{1}{2}}(x) - 1$:

$$\operatorname{sgn}(x) \begin{cases} -1 & \text{if } x < 0\,, \\ 0 & \text{if } x = 0\,, \\ 1 & \text{if } x > 0\,, \end{cases} \tag{1.26}$$

which has inversion symmetry at the origin $x_0 = 0$. The sign function is also used in combination with the Heaviside Theta function in order to specify real parts and absolute values in unified analytical expressions.[26]

The value '1' at $x = x_0 = 0$ in $\mathrm{H}_1(x)$ implies right-hand continuity. As said this convention is adopted in probability theory, in particular the cumulative distribution functions, $F_{\mathcal{Z}}(x)$ are defined to be right-hand continuous as are the integrator functions $h(x)$ in Lebesgue-Stieltjes integration (section 1.8 and leads to semi-differentiability to the right. Right-hand continuity

[26] Program packages for computer assisted calculations commonly contain several differently defined step functions. *Mathematica*, for example, uses a Heaviside Theta function with the definition (1.24): $\mathrm{H}(0)$ is undefined but $\mathrm{H}(0) - \mathrm{H}(0) = 0$ and $\mathrm{H}(0)/\mathrm{H}(0) = 1$, a *unit step* function with right-hand continuity, which is defined as $\mathrm{H}_1(x)$, and a *sign function* according to (1.26).

is the applied in conventional handling of stochastic processes. An example are semimartigales (section 3.1.3) for which the *càdlàg*-property is basic.

The behavior of step functions is easily expressed in terms of *indicator functions*, which we discuss here as another class of step functions. The indicator function of the event A in Ξ is a mapping of Ξ into 0 and 1, $\mathbf{1}_A : \Xi \to \{0,1\}$ with the properties

$$\mathbf{1}_A(x) \;=\; \begin{cases} 1 & \text{iff } x \in A \\ 0 & \text{iff } x \notin A \end{cases} \quad . \tag{1.27a}$$

Accordingly, $\mathbf{1}_A(x)$ extracts the point of the subset $A \in \Xi$ from a set Ξ that might be the entire sample set $\Xi \equiv \Omega$. For a probability space characterized by the triple $(\Omega, \Xi, \mathcal{P})$ with $\Xi \in \Pi(\Omega)$ we define an indicator random variable $\mathbf{1}_A : \Omega \to \{0,1\}$ with the properties $\mathbf{1}_A(\omega) = 1$ if $\omega \in A$, otherwise $\mathbf{1}_A(\omega) = 0$, and this yields the expectation value

$$\mathrm{E}\big(\mathbf{1}_A(\omega)\big) \;=\; \int_\Xi \mathbf{1}_A(x)\,\mathrm{d}\mathcal{P}(x) \;=\; \int_A \mathrm{d}\mathcal{P}(x) \;=\; P(A)\;, \tag{1.27b}$$

as well as the variance and covariance

$$\begin{aligned} \mathrm{var}\big(\mathbf{1}_A(\omega)\big) \;&=\; P(A)\,\big(1 - P(A)\big)\,, \text{ and} \\ \mathrm{cov}\big(\mathbf{1}_A(\omega,)\mathbf{1}_B(\omega)\big) \;&=\; P(A \cap B) \;-\; P(A)P(B). \end{aligned} \tag{1.27c}$$

We shall use indicator functions in forthcoming sections for the calculation of Lebesgue integrals (section 1.8.3) and for convenient solutions of principal value integrals by partitioning of the domain of integration (section 3.2.5).

1.6.3 Discrete probability distributions

Discrete random variables are fully characterized by either of the two probability distributions, the probability mass function (pmf) and the cumulative probability distribution (cdf). Both functions have been mentioned already and were illustrated in figures 1.7 and 1.9, respectively, and they are equivalent in the sense that essentially all observable properties can be calculated from either one of them. Because of their general importance we summarize the most important properties of discrete probability distributions.

Making use of our knowledge on probability space the probability mass function (pmf) can be formulated as a mapping from sample space into the real numbers, which returns the probability that a discrete random variable $\mathcal{Z}(\omega)$ attains exactly some value $x = x_k$. Let $\mathcal{Z}(\omega)$ be a discrete random variable on the sample space Ω, $\mathcal{Z} : \Omega \to \mathbb{R}$, and then the probability mass function is a mapping onto the unit interval, $f_\mathcal{Z} : \mathbb{R} \to [0,1]$:

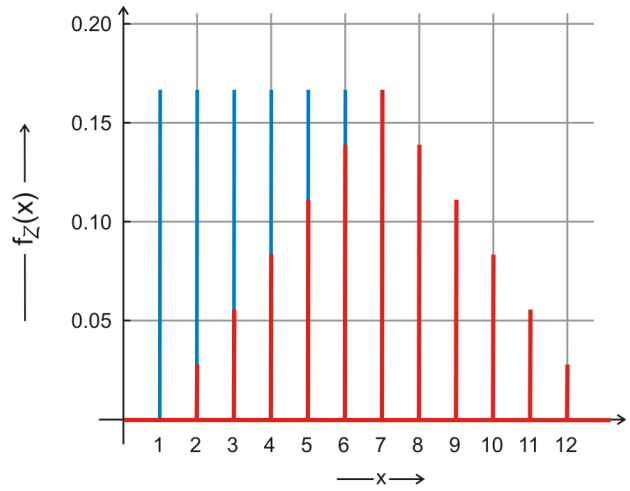


Fig. 1.11 Probability mass function of rolling fair dice. The figure shows the probability mass function (pmf), $f_{\mathcal{Z}}(x_k)$, for rolling one die or two dice simultaneously. The scores x_k are plotted on the abscissa axis. The pmf is zero everywhere on the x-axis except at a set of points, $x_k \in \{1,2,3,4,5,6\}$ for one die and $x_k \in \{2,3,4,5,6,7,8,9,10,11,12\}$ for two dice corresponding to the possible scores, $f_{\mathcal{Z}}(x_k) = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$ for one die (blue) and $f_{\mathcal{Z}}(x_k) = (1/36, 1/18, 1/12, 1/9, 5/36, 1/6, 5/36, 1/9, 1/12, 1/18, 1/36)$ (red) for two dice, respectively. In the latter case the maximal probability value is obtained for the score $x = 7$ (see also equation (1.28') and figure 1.7).

$$f_{\mathcal{Z}}(x_k) \;=\; P\big(\{\omega \in \Omega \,|\, \mathcal{Z}(\omega) = x_k\}\big) \;\text{ with }\; \sum_{k=1}^{\infty} f_{\mathcal{Z}}(x_k) = 1 \;, \tag{1.28}$$

where the probability could also be simpler expressed as $P(\mathcal{Z} = x_k)$. Sometimes it is useful to be able to treat a discrete probability distribution as if it were continuous. The function $f_{\mathcal{Z}}(x)$ is then defined for all real numbers, $x \in \mathbb{R}$ including those outside the sample set. Then we have: $f_{\mathcal{Z}}(x) = 0 \;\forall\; x \notin \mathcal{Z}(\Omega)$. A simple but straightforward representation of the probability mass function makes use of the Dirac delta-function.[27] The nonzero score values are assumed to lie exactly at the positions x_k with $k \in \mathbb{N}_{>0}$ and $p_k = P(\mathcal{Z} = x_k)$:

$$f_{\mathcal{Z}}(x) \;=\; \sum_{k=1}^{\infty} P(\mathcal{Z} = x_k)\, \delta(x - x_k) \;=\; \sum_{k=1}^{\infty} p_k\, \delta(x - x_k) \;. \tag{1.28'}$$

[27] The delta-function is no proper function but a generalized function or *distribution*. It was introduced by Paul Dirac in quantum mechanics. For more details see, for example, [393, pp.585-590] and [384, pp.38-42].

In this form the probability density function is suitable for the calculation of derived probabilities by integration (1.29').

The cumulative distribution function (cdf) of a discrete probability distribution is a step function and contains, in essence, the same information as the probability mass function. It is again a mapping from sample space into the real numbers on the unit interval, $F_{\mathcal{Z}} : \mathbb{R} \to [0, 1]$, and defined by

$$F_{\mathcal{Z}}(x) \;=\; P(\mathcal{Z} \leq x) \;\text{ with }\; \lim_{x \to -\infty} F_{\mathcal{Z}}(x) = 0 \text{ and } \lim_{x \to +\infty} F_{\mathcal{Z}}(x) = 1\,. \quad (1.29)$$

By definition the cumulative distribution functions are continuous and differentiable on the right-hand side of the steps, they cannot be integrated by conventional Riemann integration but they are Riemann-Stieltjes or Lebesgue integrable (see section 1.8). Since the integral of the Dirac delta-function is the Heaviside function we may also write

$$F_{\mathcal{Z}}(x) \;=\; \int_{-\infty}^{x} f_{\mathcal{Z}}(s)\,\mathrm{d}s \;=\; \sum_{x_k \leq x} p_k\;. \quad (1.29')$$

This integral expression is convenient because it holds for discrete and continuous probability distributions.

Special cases of importance in physics and chemistry are integer valued positive random variables, $\mathcal{Z} \in \mathbb{N}$, corresponding to a countably infinite sample space which is the set of non-negative integers: $\Omega = \mathbb{N}$ with

$$p_k \;=\; P(\mathcal{Z} = k)\,,\; k \in \mathbb{N} \;\text{ and }\; F_{\mathcal{Z}}(x) \;=\; \sum_{0 \leq k \leq x} p_k\;. \quad (1.30)$$

Such integer valued random variables will be used, for example, in master equations for modeling particle numbers or other discrete quantities in stochastic processes.

For the purpose of illustration we choose again throwing dice (figures 1.11 and 1.12). For rolling one die with s faces the pmf consists of s isolated peaks, $f_{1d}(x_k) = 1/s$ at $x_k = 1, 2, \ldots, s$ and has the value $f_{\mathcal{Z}}(x) = 0$ everywhere else ($x \neq 1, 2, \ldots, s$). Rolling two dice leads to a pmf in the form of a tent function shown in figure 1.11:

$$f_{2d}(x_k) \;=\; \begin{cases} \frac{1}{s^2}\,(k-1) & \text{for } \; k \;=\; 1, 2, \ldots, s\;, \\ \frac{1}{s^2}\,(2s+1-k) & \text{for } \; k \;=\; s+1, s+2, \ldots, 2s \end{cases}\;.$$

Here k is the score and s the number of faces of the die, which is six in case of the commonly used dice. The cumulative probability distribution function (cdf) is an example of for an ordered sum of random variables. The scores of rolling one die or two dice simultaneously are the events. The cumulative probability distribution is simply given by the sum of scores (figure 1.12):

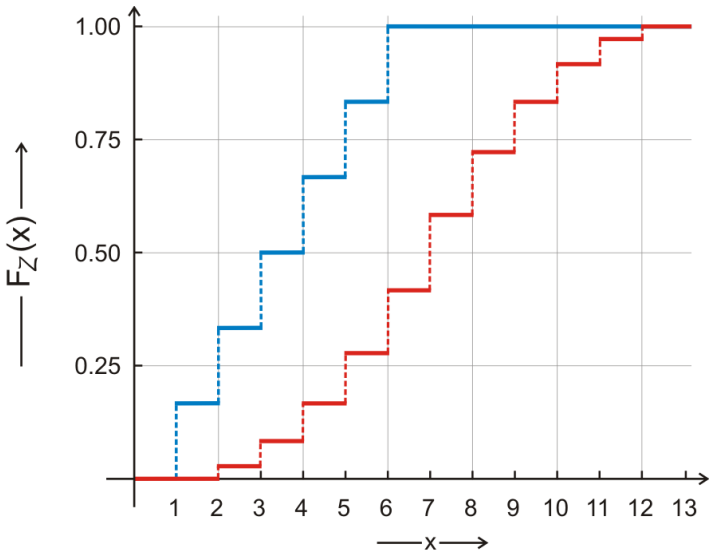


Fig. 1.12 The cumulative distribution function of rolling fair dice. The cumulative probability distribution function (cdf) is a mapping from the sample space Ω onto the unit interval $[0,1]$ of $\mathbb{R}$. It corresponds to the ordered partial sum with the ordering parameter being the score given by the stochastic variable. The example shown deals with throwing fair dice: The distribution for one die (blue) consists of six steps of equal height $p_k = 1/6$ at the scores $x_k = 1,2,\ldots,6$. The second curve (red) is the probability of throwing two dice yielding the scores $x_k = 2,3,\ldots,12$ where the weights for the individual scores are $p_k = (1/36, 1/18, 1/12, 1/9, 5/36, 1/6, 5/36, 1/9, 1/12, 1/18, 1/36)$. The two limits of a cdf are $\lim_{x\to-\infty} F_{\mathcal{Z}}(x) = 0$ and $\lim_{x\to+\infty} F_{\mathcal{Z}}(x) = 1$.

$$F_{2d}(k) \;=\; \sum_{i=2}^{k} f_{2d}(i)\,;\;\; k = 2,3,\ldots,2s\;.$$

A generalization to rolling n dice will be presented in chapter 2.6 in the discussion of the *central limit theorem*.

Finally, we generalize to sets that define the domain of a random variable on the closed interval $[a,b]$.[28] This is tantamount to restricting the sample set to these sample points, which give rise to values of the random variable on the interval:

$$\{a \le \mathcal{Z} \le b\} \;=\; \{\omega |\, a \le \mathcal{Z}(\omega) \le b\}\;,$$

[28] The notation we are applying here uses square brackets, '[··]', for closed intervals, open square brackets, ']··[', for open intervals, ']··]' and '[··[' for left-hand or right-hand half-open intervals, respectively. An alternative less common notation uses parentheses instead of open square brackets, e.g., '(··)' instead of ']··['.

and define their probabilities by $P(a \leq \mathcal{Z} \leq b)$. The set of sample points for event A, of course, need not be a closed interval, it may be an open, a half-open, an infinite intervals, or even a single point x. Then it is called a *singleton* $\{x\}$ with $P(\mathcal{Z} = x) = P(\mathcal{Z} \in \{x\})$.

For any countable, finite or countably infinite, sample spaces Ω the exact range of $\mathcal{Z}$ is just the set of the real numbers w_i:

$$W_{\mathcal{Z}} \;=\; \bigcup_{\omega \in \Omega} \{\mathcal{Z}(\omega)\} \;=\; \{w_1, w_2, \ldots, w_n, \ldots\} \text{ with } p_k \;=\; P(\mathcal{Z} = w_k)\,,\; w_k \in W_{\mathcal{Z}}.$$

As with the probability mass function (1.28') we have $P(\mathcal{Z} = x) = 0$ if $x \notin W_{\mathcal{Z}}$. Knowledge of all p_k-values is tantamount to having full information on all probabilities derivable for the random variable $\mathcal{Z}$:

$$P(a \leq \mathcal{Z} \leq b) \;=\; \sum_{a \leq w_k \leq b} p_k \quad \text{or, in general,} \quad P(\mathcal{Z} \in A) \;=\; \sum_{w_k \in A} p_k \;. \tag{1.31}$$

The cumulative distribution function (1.29) of $\mathcal{Z}$ is the special case for which A is the infinite interval $]-\infty, x]$. It fulfils several properties on intervals,

$$\begin{aligned} F_{\mathcal{Z}}(a) \;-\; F_{\mathcal{Z}}(b) \;&=\; P(\mathcal{Z} \leq b) \;-\; P(\mathcal{Z} \leq a) \;=\; P(a < \mathcal{Z} \leq b)\;, \\ P(\mathcal{Z} = x) \;&=\; \lim_{\epsilon \to 0} \big(F_{\mathcal{Z}}(x+\epsilon) \;-\; F_{\mathcal{Z}}(x-\epsilon)\big)\;, \text{ and} \\ P(a < \mathcal{Z} < b) \;&=\; \lim_{\epsilon \to 0} \big(F_{\mathcal{Z}}(b-\epsilon) \;-\; F_{\mathcal{Z}}(a+\epsilon)\big)\;, \end{aligned}$$

which can be easily verified.

1.6.4 Conditional probabilities and independence

Probabilities of events A were defined so far in relation to the entire sample space Ω, $P(A) = |A|/|\Omega| = \sum_{\omega \in A} P(\omega) \;/\; \sum_{\omega \in \Omega} P(\omega)$. Now we want to know the probability of an event A relative to a subset of sample space Ω, and we denote this set by S. This means that we attempt to calculate the proportional weight of the part of the subset A in S, which is expressed by the intersection $A \cap S$ relative to the set S, and obtain

$$\sum_{\omega \in A \cap S} P(\omega) \;\Big/\; \sum_{\omega \in S} P(\omega)\;.$$

In other words, we switch from Ω to S as the new universe and the set to be weighted are the sample points belonging to both, A and to S. It is often helpful for imagination to call the event S a *hypothesis*, which reduces the sample space from Ω to S for the definition of conditional probabilities.

The *conditional probability* measures the probability of A *relative to* S:

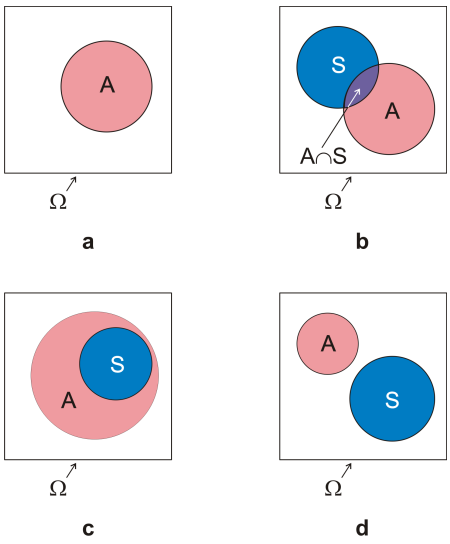


Fig. 1.13 Conditional probabilities. Conditional probabilities measure the intersection of the sets for two events, $A \cap S$ relative to the set S: $P(A|S) = |AS|/|S|$. In essence this is the same kind of weighting that defines the probabilities in sample space: $P(A) = |A|/|\Omega|$ (Part **a** shows $A \subset \Omega$ and **b** shows $A \cap S \subset S$). The two extremes are: $A \cap S = S$ and $P(A|S) = 1$ (**c**) and $A \cap S = 0$ and $P(A|S) = 0$ (**d**).

$$P(A|S) \;=\; \frac{P(A \cap S)}{P(S)} \;=\; \frac{P(AS)}{P(S)} \tag{1.32}$$

provided $P(S) \neq 0$. The conditional probability $P(A|S)$ is undefined for hypothesis of zero probability, $S = \emptyset$. Apparently, the conditional probability vanishes when the intersection is empty: $P(A|S) = 0$ if $A \cap S = AS = \emptyset$,[29] and $P(AS) = 0$. In case S is a true subset of A, $AS = S$ we have $P(A|S) = 1$ (figure 1.13).

The definition of the conditional probability implies that all general theorems on probabilities hold by the same token for conditional probabilities and, for example, we derive from equation (1.13):

$$P(A \cup B|S) \;=\; P(A|S) \;+\; P(B|S) \;-\; P(AB|S)\;. \tag{1.13'}$$

Additivity of conditional probability requires an empty intersection, $AB = \emptyset$.

Equation (1.32) is particularly useful when written in slightly different form:

$$P(AS) \;=\; P(A|S) \cdot P(S)\;, \tag{1.32'}$$

[29] From here on we shall use the short notation for the intersection, $AS \equiv A \cap S$.

which is known as the *theorem of compound probabilities* and which can be easily generalized to more events. For three events we derive [133, chap.V]

$$P(ABC) \;=\; P(A|BC)\cdot P(B|C)\cdot P(C)$$

by applying (1.32') twice – first by setting $BC \equiv S$ and then by setting $BC \equiv AS$. For n arbitrary events $A_i; i = 1,\ldots,n$ this leads to

$$P(A_1A_2\ldots A_n) \;=\; P(A_1|A_2A_3\ldots A_n)\cdot P(A_2|A_3\ldots A_n)\;\ldots\; P(A_{n-1}|A_n)\cdot P(A_n)$$

provided $P(A_2A_3\ldots A_n) > 0$. If the intersection of event sets $A_2\ldots A_n$ does not vanish, all conditional probabilities are well defined since

$$P(A_n) \geq P(A_{n-1}A_n) \geq \ldots \geq P(A_2A_3\ldots A_n) > 0\ .$$

Next we derive an equation that we shall need in chapter 3 for modeling of stochastic processes. We assume that the sample space Ω is partitioned into n disjoint sets, $\Omega = \sum_n S_n$, then we have for any set A

$$A \;=\; AS_1 \cup AS_2 \cup \ldots \cup AS_n$$

and from equation (1.32') we get

$$P(A) \;=\; \sum_n P(A|S_n)\cdot P(S_n)\ . \tag{1.33}$$

From this relation it is straightforward to derive the conditional probability

$$P(S_j|A) \;=\; \frac{P(S_j)P(A|S_j)}{\sum_n P(S_n)P(A|S_n)}$$

provided $P(A) > 0$.

The conditional probability can also be interpreted that the information on whether or not an event S has occurred changes the probability of A. Independence of events can be easily formulated in terms of conditional probabilities, since it implies that an influence of S on A does not exist and hence $P(A|S) = P(A)$ defines *stochastic independence*. Making use of equation (1.32') we define

$$P(AS) \;=\; P(A)\cdot P(S)\ , \tag{1.34}$$

and realize an important symmetry of stochastic independence: A is independent of S implies S is independent of A, and we may account for this symmetry by defining independence by stating that A and S are independent if equation (1.34) holds. We remark that the definition (1.34) is acceptable also for $P(S) = 0$ a case in which $P(A|S)$ is undefined [133, p. 125].

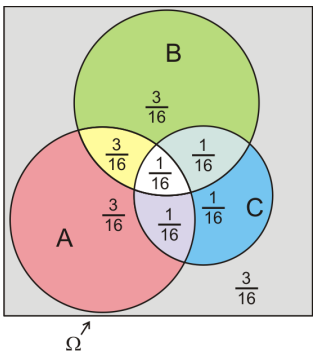


Fig. 1.14 Testing for stochastic independence of three events. The study case shown here is a example for independence of three events and corresponds to example **a** in table 1.2. The numbers in the sketch fulfill equations (1.35a) and (1.35b). The probability of the union of all three sets is given by the relation:
$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(AC) + P(ABC)$,
and by addition of the remainder one verifies $P(\Omega) = 1$.

The case of more than two events needs some care and we take three events A, B, C as an example. So far we were dealing with pairwise independence and accordingly we have

$$P(AB) = P(A) \cdot P(B)\,,\; P(BC) = P(B) \cdot P(C)\,,\; P(CA) = P(C) \cdot P(A)\,. \quad (1.35a)$$

Pairwise independence, however, does not necessarily imply that

$$P(ABC) \;=\; P(A) \cdot P(B) \cdot P(C) \quad (1.35b)$$

holds. In addition, examples were constructed where the last equation is fulfilled but nevertheless the sets are not pairwise independent [160].

Independence or lack of independence of three sets can be easily visualized by means of weighted Venn diagrams. In figure 1.14 we show a case where independence of the three sets A, B, and C is easily tested. Although cases of pairwise independence but lacking mutual independence of three events are not common they can be found in general: Case **f** in figure 1.4 allows for straightforward construction of examples with pairwise independence but $P(ABC) = 0$. Eventually, we present also an opposite example, which is attributed to Sergei Bernstein [133, p. 127]: The six permutations of the three letters a, b and c together with the three triples (aaa), (bbb), (ccc) constitute the sample space and a probability $P = \frac{1}{9}$ is attributed to each sample point. Now we define three events A_1, A_2 and A_3 according to the appearance of the letter a at the first, second or third place:

$$A_1 \;=\; \{aaa, abc, acb\}\,,\; A_2 \;=\; \{aaa, bac, cab\}\,,\; A_3 \;=\; \{aaa, bca, cba\}\,.$$

Table 1.2 Testing for stochastic independence of three events. We show three examples: Case **a** fulfils equations (1.35a) and (1.35b), and represents a case of mutual independence (figure 1.14), case **b** fulfils only equation (1.35a) and not equation (1.35b), and is an example of pairwise independent but not mutually independent events, and case **c** is an especially constructed example for fulfilment of equation (1.35b) by three sets that are pairwise independent. Deviations from equations (1.35a) and (1.35b) are shown in boldface numbers.

	Probabilities P						
	Singles			Pairs			Tripel
	A	B	C	AB	BC	CA	ABC
a	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$
b	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\mathbf{\frac{1}{10}}$
c	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{2}$	$\mathbf{\frac{1}{10}}$	$\mathbf{\frac{6}{25}}$	$\mathbf{\frac{7}{50}}$	$\frac{1}{25}$

Every event has a probability $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$ and the three events are pairwise independent because

$$P(A_1A_2) \;=\; P(A_2A_3) \;=\; P(A_3A_1) \;=\; \frac{1}{9}\,,$$

but they are not mutually independent because $P(A_1A_2A_3) = \frac{1}{9}$ instead of $\frac{1}{27}$ as required by equation (1.35b). In this case it is easy to detect the cause of the mutual dependence: The occurrence of two events implies the occurrence of the third and therefore we have $P(A_1A_2) = P(A_2A_3) = P(A_3A_1) = P(A_1A_2A_3)$. Table 1.2 finally presents numerical examples for all three cases.

Generalization to n events is straightforward [133, p. 128]: The events $A_1\,,\, A_2\,,\, \ldots\,,\, A_n$ are mutually independent if the multiplication rules apply for all combinations $1 \le i < j < k < \ldots \le n$ and hence we have $2^n - n - 1$ conditions,

$$\begin{aligned} P(A_iA_j) \;&=\; P(A_i)\cdot P(A_j) \\ P(A_iA_jA_k) \;&=\; P(A_i)\cdot P(A_j)\cdot P(A_k) \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ P(A_1A_2\ldots A_n) \;&=\; P(A_1)\cdot P(A_2)\cdot \ldots \cdot P(A_n)\,, \end{aligned} \tag{1.36}$$

which have to be satisfied.[30]

Two or more random variables,[31] for example $\mathcal{X}$ and $\mathcal{Y}$, can be subsumed in a *random vector* $\vec{\mathcal{V}} = (\mathcal{X}, \mathcal{Y})$, which is expressed by the *joint probability*

$$P(\mathcal{X} = x_i, \mathcal{Y} = x_j) \;=\; p\,(x_i, y_j)\;. \tag{1.37}$$

The random vector $\vec{\mathcal{V}}$ is fully determined by the *joint probability mass function*

$$\begin{aligned} f_{\vec{\mathcal{V}}}(x,y) \;&=\; P(\mathcal{X} = x, \mathcal{Y} = y) \;=\; P(\mathcal{X} = x \;\vee\; \mathcal{Y} = y) \;= \\ &=\; P(\mathcal{Y} = y \,|\, \mathcal{X} = x) \cdot P(\mathcal{X} = x) \;= \\ &=\; P(\mathcal{X} = x \,|\, \mathcal{Y} = y) \cdot P(\mathcal{Y} = y)\;. \end{aligned} \tag{1.38}$$

This density constitutes the probabilistic basis of the random vector $\vec{V}$. It is straightforward to define a *cumulative probability distribution* in analogy to the single variable case

$$F_{\vec{\mathcal{V}}}(x,y) \;=\; P(\mathcal{X} \le x, \mathcal{Y} \le y)\;. \tag{1.39}$$

In principle either of the two probability functions contain the complete information on both variables but depending on the specific situation either the pmf or the cdf may be more efficient.

Often no detailed information is required on one particular random variable. Then, by summation over one variable of the vector $\vec{\mathcal{V}}$ we obtain the probabilities for the corresponding *marginal distribution*,

$$\begin{aligned} P(\mathcal{X} = x_i) \;&=\; \sum_{y_j} p\,(x_i, y_j) \;=\; p\,(x_i, *) \;\text{ and} \\ P(\mathcal{Y} = y_j) \;&=\; \sum_{x_i} p\,(x_i, y_j) \;=\; p\,(*, y_j)\;, \end{aligned} \tag{1.40}$$

of $\mathcal{X}$ and $\mathcal{Y}$, respectively.

Independence of random variables will be a highly relevant problem in the forthcoming chapters. Countably-valued random variables $\mathcal{X}_1, \ldots, \mathcal{X}_n$ are defined to be *independent* if and only if for any combination $x_1, \ldots, x_n$ of real numbers the joint probabilities can be factorized:

$$P(\mathcal{X}_1 = x_1, \ldots, \mathcal{X}_n = x_n) \;=\; P(\mathcal{X}_1 = x_1) \cdot \ldots \cdot P(\mathcal{X}_n = x_n)\;. \tag{1.41}$$

---

[30] The number of conditions consists of $\binom{n}{2}$ equations in the first line, $\binom{n}{3}$ equations in the second line, and so on, down to $\binom{n}{n} = 1$ in the last line. The summation yields $\sum_{i=2}^{n} \binom{n}{i} = (1+1)^n - \binom{n}{1} - \binom{n}{0} = 2^n - n - 1$.

[31] For simplicity we restrict ourselves to the two variable case here. The extension to any finite number of variables is straightforward.

An extension of equation (1.41) replaces the single values x_i by arbitrary sets S_i

$$P(\mathcal{X}_1 \in S_1, \ldots, \mathcal{X}_n \in S_n) \;=\; P(\mathcal{X}_1 \in S_1) \cdot \ldots \cdot P(\mathcal{X}_n \in S_n) \;.$$

In order to proof this extension we sum over all points belonging to the sets $S_1, \ldots, S_n$:

$$\begin{aligned} &\sum_{x_1 \in S_1} \cdots \sum_{x_n \in S_n} P(\mathcal{X}_1 = x_1, \ldots, \mathcal{X}_n = x_n) \;= \\ = &\sum_{x_1 \in S_1} \cdots \sum_{x_n \in S_n} P(\mathcal{X}_1 \in S_1) \cdot \ldots \cdot P(\mathcal{X}_n \in S_n) \;= \\ = &\left(\sum_{x_1 \in S_1} P(\mathcal{X}_1 \in S_1) \right) \cdot \ldots \cdot \left(\sum_{x_n \in S_n} P(\mathcal{X}_n \in S_n) \right) \;, \end{aligned}$$

which is equal to the right hand side of the equation to be proven. □

Since the factorization is fulfilled for arbitrary sets $S_1, \ldots S_n$ it holds also for all subsets of $(\mathcal{X}_1 \ldots \mathcal{X}_n)$ and accordingly the events

$$\{\mathcal{X}_1 \in S_1\}, \ldots, \{\mathcal{X}_n \in S_n\}$$

are also independent. It can also be verified that for arbitrary real-valued functions $\varphi_1, \ldots, \varphi_n$ on $]-\infty, +\infty[$ the random variables $\varphi_1(\mathcal{X}_1), \ldots, \varphi_n(\mathcal{X}_n)$ are independent too.

Independence can also be extended in straightforward manner to the joint distribution function of the random vector $\vec{\mathcal{V}} = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$

$$F_{\vec{\mathcal{V}}}(x_1, \ldots, x_n) \;=\; F_{\mathcal{X}_1}(x_1) \cdot \ldots \cdot F_{\mathcal{X}_n}(x_n) \;,$$

where the $F_{\mathcal{X}_j}$'s are the marginal distributions of the $\mathcal{X}_j$'s , $1 \leq j \leq n$. Thus, the marginal distributions determine the joint distribution in case of independence of the random variables.

1.7 Probability measure on uncountable sample spaces[32]

In the previous sections we were dealing with countable, finite or infinite, sample spaces where classical probability theory would have worked as well as the set theoretic approach. A new situation arises when the sample space Ω is uncountable (see, e.g., figure 1.5) and this is very common, for example, for continuous variables defined on non-zero, open or closed segments of the real line, $]a,b[$, $]a,b]$, $[a,b[$, or $[a,b]$ for $a<b$, respectively. The most straightforward way to illustrate the definition of a measure on an uncountable sample space is to assign length (m), area (m^2), volume (m^3), or generalized volume (m^n) to the uncountable set. In order to illustrate the problem we may ask a very natural question: Does every arbitrary proper subset of the real line, $-\infty<x<+\infty$, have a length? It seems trivial to assign length 1 to the interval $[0,1]$ and length $b-a$ to the interval $[a,b]$ with $a\leq b$. Often the mass of a homogeneous object is easier to imagine than the volume, and we assign mass to sets in the sense of bars of uniform density. For example, we attribute a bar of length 1 that has mass 1 to $[0,1]$, and accordingly, a bar of mass $b-a$ to $[a,b]$, two bars corresponding to the set $[0,1]\cup[3,5]$ together have mass 3, etc. The question now is: *What is the mass of the set of the rational numbers* $\mathbb{Q}$ *given the mass of the interval* $[0,1]$ *is one?* Since the rational numbers are *dense* in the real numbers,[33] any nonnegative value for the mass of the rational numbers may appear acceptable. The real numbers, however, are uncountable and so are the irrational numbers, $\mathbb{R}\backslash\mathbb{Q}$. Assigning mass $b-a$ to the interval $[a,b]$ leaves no weight for the rational numbers and indeed the rational numbers $\mathbb{Q}$ have measure zero like any other set of countably many objects.

In this section we have to be more precise and call the measure we are introducing here a *Lebesgue measure*: The rational numbers have Lebesgue measure zero, $\lambda(\mathbb{Q})=0$. In the forthcoming text we shall see that the Lebesgue measure indeed assigns precisely the values given above to intervals on the real axis: $\lambda([0,1])=1$ or $\lambda([a,b])=b-a$. The real line $\mathbb{R}$ allows for the definition of a Borel measure, which assigns $\mu([a,b])=b-a$ for every interval $[a,b]$ too. The Borel measure is defined on the σ-algebra (see section 1.7.2)[34] of the Borel sets $\mathcal{B}(\mathbb{R})$ and this is the smallest σ-algebra that contains the open intervals of $\mathbb{R}$. In practice, however, the Borel measure is not the most useful measure defined on the σ-algebra of Borel sets, because in contrast to

[32] This section can be skipped by readers who are willing to except the fact that all uncountable sample spaces needed in the forthcoming discussions are measurable notwithstanding the existence of non-measurable sets.

[33] A subset D of real numbers is said to be *dense* in $\mathbb{R}$ if every arbitrarily small interval $]a,b[$ with $a<b$ contains at least one element of D. Accordingly, the set of rational numbers $\mathbb{Q}$ as well as the set of irrational numbers $\mathbb{R}\backslash\mathbb{Q}$ are dense in $\mathbb{R}$.

[34] For the time being it is sufficient to know that a σ-algebra on a set Ξ is a collection Σ of subsets $A\in\Xi$, which fulfill certain properties among them σ-additivity (see section 1.5.1).

the Borel measure the Lebesgue-measure on Borel sets is a *complete measure*. The difference between Borel and Lebesgue measures concerns the handling of *null sets* that are sets N of measure zero, $\mu(N) = 0$, for which every measurable subset $A \subset N$ has measure $\mu(A) = 0$. A complete measure refers to a complete measure space in which every subset of every null set is measurable with measure zero. Indeed, the Lebesgue measure λ is an extension of the Borel measure μ, every Borel-measurable set E is also a Lebesgue-measurable set, and the two measures coincide on Borel sets E: $\lambda(E) = \mu(E)$.

In the case of a countable sample space the powerset $\Pi(\Omega)$ is the set of all subsets of the sample space Ω and contains the results of all set theoretic operations of section 1.4. Although is would seem straightforward to proceed in the same way for uncountable sample spaces Ω, it turns out that the powerset $\Pi(\Omega)$ is too large, because it contains uncountably many subsets. For the development of a measure for uncountable sample spaces we recall the three indispensable properties of probability measures $\mu : \Xi \to [0, \infty[$ with Ξ being a measurable collection of events A:

(P) nonnegativity, $\mu(A) \geq 0 \; \forall \; A \in \Xi$,
(N) normalization, $P(\Omega) = 1$, and
(A) additivity, $\mu(A) + \mu(B) = \mu(A \cup B)$ provided $P(A \cap B) = \emptyset$.

In essence, the task is now to find measures for uncountable sets that are derived from collections of subsets, whose cardinality is $\aleph_0$ – infinite but countable. Problems concerning measurability arise from the impossibility to assign a probability to every subset of Ω. In other words, there might be sets to which no measure – no length, no mass, etc. – can be assigned. To derive the concept of measurable sets in full rigorousity is highly demanding and it requires advanced mathematical techniques, in particular sufficient knowledge of measure theory. For the probability concept we are using here, however, the simplest bridge from countability to uncountability is sufficient and we need only derive a measure for a certain family of sets called *Borel sets*, $\mathcal{B} \subset \Omega$. For this goal the introduction of σ-additivity (1.15) and Lebesgue measure $\lambda(A)$ is sufficient, and as said, σ-additivity comes close to the assignment of additive mass in the above given example. Still unanswered so far is the question whether unmeasurable sets do exist at all.

1.7.1 Existence of non-measurable sets

A general proof of this conjecture is difficult but Giuseppe Vitali [449, 450] provided a proof by means of contradiction : No mapping $P\colon \Pi(\Omega) \to [0, 1]$ exists, which fulfils all three indispensable properties for probabilities. In particular he showed this for the infinitely repeated coin flip, $\Omega = \{0, 1\}^{\mathbb{N}}$ [161, p. 9,10]:

(N) normalization: $P(\Omega) = 1$,
(A) σ-additivity: for pairwise disjoint events $A_1, A_2, \ldots \subset \Omega$ holds

$$P\left(\bigcup_{i\geq 1} A_i\right) \;=\; \sum_{i\geq 1} P\left(A_i\right) \;, \quad \text{and}$$

(I) invariance: For all $A \subset \Omega$ and $k \geq 1$ holds $P(T_k A) = P(A)$, where T_k is an operator that inverts the outcome of the k-th toss.

In particular, the sample points of Ω are $\omega = (\omega_1, \omega_2, \ldots)$ and the operator is defined by

$$T_k : \; \omega = (\omega_1, \ldots, \omega_{k-1}, \omega_k, \omega_{k+1} \ldots) \to (\omega_1, \ldots, \omega_{k-1}, 1-\omega_k, \omega_{k+1} \ldots),$$

and $T_k A = \{T_k(\omega) : \omega \in A\}$ is the image of A under the operation T_k, which defines a mapping of Ω onto itself. The first two conditions, (N) and (A), are the criteria for probability measures and the invariance condition (I) is specific for coin flipping and encapsulates the properties derived from the uniform distribution, $\mathcal{U}_\Omega$, which for the single coin toss is $P(\omega_k) = P(1-\omega_k) = \frac{1}{2}$.

Proof. In order to proof the conjecture of incompatibility with all three conditions we define an equivalence relation '$\sim$' in Ω: $\omega \sim \omega'$ iff $\omega_k = \omega'_k$ for all sufficiently long sequences with $n \geq k$, and the elements of the equivalence class are the sequences which have the same digit at position k. According to the axiom of choice[35] there exists a set $A \subset \Omega$, which contains exactly one element of each equivalence class. We define $\mathcal{S} = \{S \subset \mathbb{N} : |S| < \infty\}$ the set containing all finite subsets of $\mathbb{N}$. Since $\mathcal{S}$ is the union of a countable number of finite sets, $\{S \subset \mathbb{N} : \max S = m\}$ with $m \in \mathbb{N}$, $\mathcal{S}$ is countable too. For $S = \{k_1, \ldots, k_n\} \in \mathcal{S}$ we define $T_S = \prod_{k_i \in S} T_{k_i} = T_{k_1} \circ \ldots \circ T_{k_n}$ the simultaneous flip of the digits in S. Then we have:

(i) $\Omega = \bigcup_{S\in\mathcal{S}} T_S A$ since for every sequence $\omega \in \Omega$ there exists an $\omega' \in A$ with $\omega \sim \omega'$, and accordingly an $S \in \mathcal{S}$ such that $\omega = T_S\, \omega' \in T_S\, A$,

(ii) the sets $(T_S A)_{S\in\mathcal{S}}$ are pairwise disjoint: If $T_S A \cup T_{S'} A \neq \emptyset$ were true for $S, S' \in \mathcal{S}$ then there existed an $\omega, \omega' \in A$ with $T_S\, \omega = T_{S'}\, \omega'$ and accordingly $\omega \sim T_S\, \omega = T_S\, \omega \sim \omega'$. By definition of A we had $\omega = \omega'$ and hence $S = S'$.

Applying the properties (N), (A), and (I) of the probability P we find

$$1 \;=\; P(\Omega) \;=\; \sum_{S\in\mathcal{S}} P(T_S A) \;=\; \sum_{S\in\mathcal{S}} P(A) \;. \tag{1.42}$$

Equation (1.42) cannot be fulfilled for infinitely large series of coin tosses, since all values $P(A)$ or $P(T_S A)$ are the same and infinite summation by σ-additivity (A) is tantamount to an infinite sum of the same number, which yields either 0 or ∞ but never 1 as required to fulfil (N). □

[35] Axiom of choice: Suppose that $A_\theta : \theta \in \Theta$ is a decomposition of Ω into nonempty sets. The axiom of choice guarantees that there exists at least one set C, which contains exactly one point from each A_θ: $C \cap A_\theta$ is a singleton for each θ in Θ (see [41, p. 572] and [95]).

It is straightforward to show that the set of all binary strings with countably infinite length, $B = \{0,1\}^{\mathbb{N}}$, is bijective[36] to the unit interval $[0,1]$. A more or less explicit bijection $f : B \leftrightarrow [0,1]$ can be obtained by defining an auxiliary function

$$g(s) \doteq \sum_{k=1}^{\infty} \frac{s_k}{2^k} .$$

which interprets a binary string $s = (s_1, s_2, \ldots) \in B$ as an infinite binary fraction

$$\frac{s_1}{2} + \frac{s_2}{4} + \ldots .$$

The function $g(s)$ maps B only *almost* bijectively onto $[0,1]$, because the all dyadic rationals in $]0,1[$ have two preimages each, for example

$$g(1,0,0,0,\ldots) = \frac{1}{2} = \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \ldots = g(0,1,1,1,\ldots) .$$

In order to fix this problem we reorder the rationals:

$$\big(q_n\big)_{n\geq 1} = \left(\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}, \frac{1}{16}, \ldots\right) ,$$

and find for the bijection

$$f(s) \doteq \begin{cases} q_{2n-1} & \text{if } g(s) = q_n \text{ , and } s_k = 1 \text{ for almost all } k \text{ ,} \\ q_{2n} & \text{if } g(s) = q_n \text{ , and } s_k = 0 \text{ for almost all } k \text{ ,} \\ g(s) & \text{otherwise .} \end{cases} \tag{1.43}$$

Hence Vitali's theorem applies as well to the unit interval $[0,1]$, where we are also dealing with an uncountable number of non-measurable sets. For other more detailed proofs of Vitali's theorem see, e.g., [41, p. 47].

The proof of Vitali's theorem shows the existence of non-measurable subsets within the real numbers called Vitali sets – precisely it provides evidence for subsets of the real numbers that are not Lebesgue measurable (see next subsection 1.7.2). The problem to be solved now is a rigorous reduction of the powerset to an event system Ξ such that the subsets causing the lack of countability are left aside (figure 1.15).

1.7.2 Borel σ-algebra and Lebesgue measure

In figure 1.15 we consider the three levels of sets in set theory that are relevant for our construction of an event system Ξ. The objects on the lowest level

[36] A *bijection* or bijective function implies a one-to-one correspondence between the elements of two sets.

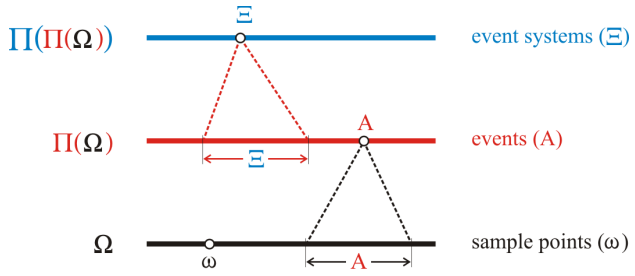


Fig. 1.15 Conceptual levels of sets in probability theory. The lowest level is the sample space Ω (black), it contains the sample points or individual results ω as elements, and events A are subsets of Ω: $\omega \in \Omega$ and $A \subset \Omega$. The next higher level is the powerset $\Pi(\Omega)$ (red). Events A are its elements and event systems Ξ constitute its subsets: $A \in \Pi(\Omega)$ and $\Xi \subset \Pi(\Omega)$. The highest level finally is the power powerset $\Pi\big(\Pi(\Omega)\big)$ that houses event systems Ξ as elements: $\Xi \in \Pi\big(\Pi(\Omega)\big)$ (blue) (Drawn after [161, p. 11]).

are the sample points corresponding to individual results, $\omega \in \Omega$. The next higher level is the powerset $\Pi(\Omega)$ housing the events $A \in \Pi(\Omega)$. The elements of the powerset are subsets of the sample space, $A \subset \Omega$. To illustrate the role of event systems Ξ we need a still higher level, the powerset of the powerset, $\Pi\big(\Pi(\Omega)\big)$: Event systems Ξ are elements of the power powerset, $\Xi \in \Pi\big(\Pi(\Omega)\big)$ and subsets of the powerset, $\Xi \subset \Pi(\Omega)$.[37]

The minimal requirements for an *event system* Ξ are summarized in the following definition of a *σ-algebra* on Ω with $\Omega \neq \emptyset$ and $\Xi \subset \Pi(\Omega)$:

Condition (1): $\Omega \in \Xi$,
Condition (2): $A \in \Xi \implies A^c \doteq \Omega \backslash A \in \Xi$, and
Condition (3): $A_1, A_2, \ldots \in \Xi \implies \bigcup_{i \geq 1} A_i \in \Xi$.

Condition (2) requires the existence of a complement A^c for every subset $A \in \Xi$, defines the logical negation as expressed by the difference between the entire sample space and the event A, and condition (3), the necessity of σ-additivity represents the logical *or* operation. The pair (Ω, Ξ) is called an *event space* and represents a *measurable space* here. From the three properties (1) to (3) follow other properties: The intersection, for example, is the complement of the union of the complements $A \cap B = (A^c \cup B^c)^c \in \Xi$, and the argument is easily extended to the intersection of countable many subsets of Ξ that belongs to Ξ as well. Thus, a σ-algebra is closed under the operations '^{c}', '$\cup$' and '$\cap$'.[38] Trivial examples of σ-algebras are $\{\emptyset, \Omega\}$, $\{\emptyset, A, A^c, \Omega\}$ or

[37] Recalling the situation in the case of countability we were choosing the entire power set $\Pi(\Omega)$ as reference instead of a smaller event system Ξ.

[38] A family of sets is called closed under an operation if the operation can be applied a countable number of times without producing a set that lies outside the family.

the family of all subsets. The Borel σ-algebra on Ω is the smallest σ-algebra, which contains all open sets or equivalently, all closed sets.

A construction principle for σ-algebras starts out from some event system $\mathcal{G} \subset \Pi(\Omega)$ (for $\Omega \neq \emptyset$) that is sufficiently small and otherwise arbitrary. Then, there exists exactly one smallest σ-algebra $\Xi = \sigma(\mathcal{G})$ in Ω with $\Xi \supset \mathcal{G}$, and we call Ξ the σ-algebra induced by $\mathcal{G}$. In other words, $\mathcal{G}$ is the generator of Ξ. Here are three important examples:

(i) the powerset with Ω being countable where $\mathcal{G} = \big\{\{\omega\} : \omega \in \Omega\big\}$ is the system of the subsets of Ω containing a single element, the σ-algebra $\sigma(\mathcal{G}) = \Pi(\Omega)$, each $A \in \Pi(\Omega)$ is countable, and $A = \bigcup_{\omega\in A}\{\omega\} \in \sigma(\mathcal{G})$ (These are the countable sample spaces as discussed in section 1.5),

(ii) the Borel σ-algebra $\mathcal{B}$ containing all open or all closed sets in one dimension (This is the uncountable sample space of real numbers $\Omega = \mathbb{R}$, see below), and

(iii) the product σ-algebra for sample spaces Ω that are Cartesian products of sets $\mathcal{E}_k$, $\Omega = \prod_{k\in\mathcal{I}} \mathcal{E}_k$ where $\mathcal{I}$ is a complete set of indices with $\mathcal{I} \neq \emptyset$ (These are the Cartesian product sample spaces of vectors with real components in n dimensions, $\Omega = \mathbb{R}^n$).

All three cases are required for the understanding of probability measures: (i) The powerset provides the frame for discrete sample spaces, (ii) the Borel σ-algebra (see below) sets the stage for one-dimensional continuous sample spaces, and (iii) the product Borel σ-algebra represents the natural extension from one dimension to the n-dimensional Cartesian space (see below).

The Borel σ-algebra[39] is constructed with the help of a generator representing the set of all compact intervals in one-dimensional Cartesian space, $\Omega = \mathbb{R}$, which have rational endpoints,

$$\mathcal{G} \;=\; \{[a, b] : a < b; (a, b) \in \mathbb{Q}\} \tag{1.44a}$$

where $\mathbb{Q}$ is the set of all rational numbers.[40] The σ-algebra induced by this generator is denoted as the Borel σ-algebra, $\mathcal{B} \doteq \sigma(\mathcal{G})$ on $\mathbb{R}$ and each $A \in \mathcal{B}$ is a Borel set.

The extension to n dimensions is straightforward and we shall consider from now on the n-dimensional case with $\Omega \in \mathbb{R}$ as the special case with $n = 1$. Another special case is our conventional physical space with $n = 3$. The generator $\mathcal{G}$ is now the set of all compact cuboids in n-dimensional Cartesian space, $\Omega = \mathbb{R}^n$, which have rational corners:

$$\mathcal{G} \;=\; \left\{ \prod_{k=1}^{n} [a_k, b_k] : a_k < b_k; a_k, b_k \in \mathbb{Q} \right\} . \tag{1.44b}$$

[39] Borel σ-algebras are frequently called Borel fields.

[40] The restriction to rational corners is the trick that makes the event system Ξ tractable in comparison of the power set, which we have shown to be too large for the definition of a Lebesgue measure.

The σ-algebra induced by this generator is called a Borel σ-algebra in n dimensions, $\mathcal{B}^{(n)} \doteq \sigma(\mathcal{G})$ on $\mathbb{R}^n$ and each $A \in \mathcal{B}^{(n)}$ is a Borel set. Then, $\mathcal{B}_k$ is a Borel σ-algebra on the subspace $\mathcal{E}_k$ with $X_k : \Omega \to \mathcal{E}_k$ being the projection onto the k-th coordinate and the generator

$$\mathcal{G} = \{X_k^{-1} A_k : k \in \mathcal{I}, A_k \in \mathcal{B}_k\}$$

is the system of all sets in Ω, which are determined by an event on coordinate k. $\mathcal{B}^{(n)} = \bigotimes_{k \in I} \mathcal{B}_k \doteq \sigma(\mathcal{G})$ is called the product σ-algebra of the sets $\mathcal{B}_k$ on Ω. In the important case of equivalent Cartesian coordinates, $\mathcal{E}_k = \mathcal{E}$ and $\mathcal{B}_k = \mathcal{B}$ for all $k \in \mathcal{I}$, a short-hand notion is common. The Borel σ-algebra on $\mathbb{R}^n$ is represented by the n-dimensional product σ-algebra of the Borel σ-algebra $\mathcal{B}$ on $\mathbb{R}$: $\mathcal{B}^n$ on $\mathbb{R}^n$.[41] .

A Borel σ-algebra is characterized by five properties, which are helpful for visualizing its enormous size:

(i) Each open set ']..[' $A \subset \mathbb{R}^n$ is Borelian. Every $\omega \in A$ has a neighborhood $Q \in \mathcal{G}$ with $Q \subset A$ and therefore we have

$$A = \bigcup_{Q \in \mathcal{G},\, Q \subset A} Q$$

representing a union of countably many sets in $\mathcal{B}^n$, which follows from condition (3) of σ-algebras.

(ii) Each closed set '[..]' $A \subset \mathbb{R}^n$ is Borelian since A^c is open and Borelian according to item (i).

(iii) The σ-algebra $\mathcal{B}^n$ cannot be described in a constructive way, because is consists of much more than the union of cuboids and their complements. In order to create $\mathcal{B}^n$ the operation of adding complements and countable unions has to be repeated as often as there are countable ordinal numbers (and this leads to an uncountable number of times [40, pp.24, 29]). For practical purposes it is sufficient to memorize that $\mathcal{B}^n$ covers almost all sets in $\mathbb{R}^n$ – but not all of them.

(iv) The Borel σ-algebra $\mathcal{B}$ on $\mathbb{R}$ is generated not only by the system of compact sets (1.44) but also by the system of closed left-hand open infinite intervals:

$$\tilde{\mathcal{G}} = \{]-\infty, c]; c \in \mathbb{R}\} \ . \tag{1.44c}$$

By analogy $\mathcal{B}$ is also generated from all open left-unbounded, from all closed, and from all open right-unbounded intervals.

(v) The event system $\mathcal{B}_\Omega^n = \{A \cap \Omega : A \in \mathcal{B}^n\}$ on $\Omega \subset \mathbb{R}^n$, $\Omega \neq \emptyset$ represents a σ-algebra on ω, which is denoted as the Borel σ-algebra on Ω.

[41] For $n = 1$ one commonly writes $\mathcal{B}$ instead of $\mathcal{B}^1$, or $\mathcal{B}^n = \mathcal{B}^{\otimes n}$.

Item (iv) follows from condition (2), which requires $\tilde{\mathcal{G}} \subset \mathcal{B}$ and – because of minimality of $\sigma(\tilde{\mathcal{G}})$ – $\sigma(\tilde{\mathcal{G}}) \subset \mathcal{B}$ too. Alternatively, $\sigma(\tilde{\mathcal{G}})$ contains all left-open intervals since $]a,b] =]-\infty,b] \setminus]-\infty,a]$ and also all compact or closed intervals since $[a,b] = \bigcap_{n\geq 1}]a-\frac{1}{n},b]$ and accordingly also the σ-algebra $\mathcal{B}$ generated from these intervals (1.44a). All intervals discussed in items (i) to (iv) are Lebesgue measurable while other sets are not.

The Lebesgue measure is the conventional mean of assigning lengths, areas, and volumes to subsets of three-dimensional Euclidean space and in formal Cartesian spaces to objects with higher dimensional volumes. Sets to which generalized volumes[42] can be assigned are called Lebesgue measurable and the measure or the volume of such a set A is denoted by $\lambda(A)$. The Lebesgue measure on $\mathbb{R}^n$ has the following properties:

(1) If A is a Lebesgue measurable set, then $\lambda(A) \geq 0$.
(2) If A is a Cartesian product of intervals, $I_1 \otimes I_2 \otimes \ldots \otimes I_n$, then A is Lebesgue measurable and $\lambda(A) = |I_1| \cdot |I_2| \cdot \ldots \cdot |I_n|$.
(3) If A is Lebesgue measurable, its complement A^c is so too.
(4) If A is a disjoint union of countably many disjoint Lebesgue measurable sets, $A = \bigcup_k A_k$, then A is Lebesgue measurable and $\lambda(A) = \sum_k \lambda(A_k)$.
(5) If A and B are Lebesgue measurable and $A \subset B$, then $\lambda(A) \leq \lambda(B)$.
(6) Countable unions and countable intersections of Lebesgue measurable sets are Lebesgue measurable.[43]
(7) If A is an open or closed subset or Borel set of $\mathbb{R}^n$, then A is Lebesgue measurable.
(8) The Lebesgue measure is strictly positive on non-empty open sets, and so its support is the entire $\mathbb{R}^n$.
(9) If A is a Lebesgue measurable set with $\lambda(A) = 0$, called a null set, then every subset of A is also a null set, and every subset of A is measurable.
(10) If A is Lebesgue measurable and $\mathbf{r}$ is an element of $\mathbb{R}^n$, then the translation of A by $\mathbf{r}$ that is defined by $A + \mathbf{r} = \{\mathbf{a} + \mathbf{r} | \mathbf{a} \in A\}$ is also Lebesgue measurable and has the same measure as A.
(11) If A is Lebesgue measurable and $\delta > 0$, then the dilation of A by δ defined by $\delta A = \{\delta \mathbf{r} | \mathbf{r} \in A\}$ is also Lebesgue measurable and has measure $\delta^n \lambda(A)$.
(12) In generalization of items (10) and (11), if L is a linear transformation and A is a measurable subset of $\mathbb{R}^n$, then $T(A)$ is also measurable and has the measure $\lambda = |\det(T)|\, \lambda(A)$.

---

[42] We generalize volume here to arbitrary dimension n: The *generalized volume* for $n = 1$ is a length, for $n = 2$ an area, for $n = 3$ a (conventional) volume and for arbitrary dimension n a cuboid in n-dimensional space.

[43] This is not a consequence of items (3) and (4): A family of sets, which is closed under complements and countable disjoint unions, need not be closed under countable non-disjoint unions. Consider, for example, the set

$$\big\{\emptyset, \{1,2\}, \{1,3\}, \{2,4\}, \{3,4\}, \{1,2,3,4\}\big\}.$$

All twelve items listed above can be succinctly summarized in one lemma:

The Lebesgue measurable sets form a σ-algebra on $\mathbb{R}^n$ containing all products of intervals, and λ is the unique complete translation-invariant measure on that σ-algebra with

$$\lambda\big([0,1] \otimes [0,1] \otimes \ldots \otimes [0,1]\big) = 1.$$

We conclude this section on Borel σ-algebra and Lebesgue measure by mentioning a few characteristic and illustrative examples:

- Any closed interval $[a,b]$ of real numbers is Lebesgue measurable, and its Lebesgue measure is the length $b-a$. The open interval $]a,b[$ has the same measure, since the difference between the two sets consists of the two endpoint a and b only and has measure zero.
- Any Cartesian product of intervals $[a,b]$ and $[c,d]$ is Lebesgue measurable and its Lebesgue measure is $(b-a)\cdot(d-c)$ the area of the corresponding rectangle.
- The Lebesgue measure of the set of rational numbers in an interval of the line is zero, although this set is dense in the interval.
- The Cantor set[44] is an example of an uncountable set that has Lebesgue measure zero.
- Vitali sets are examples of sets that are not measurable with respect to the Lebesgue measure.

In the forthcoming sections we make use of the fact that the continuous sets on the real axes become countable and Lebesgue measurable if rational numbers are chosen as beginnings and end points of intervals. Hence, we can work with real numbers with almost no restriction for practical purposes.

[44] The Cantor set is generated from the interval $[0,1]$ through consecutively taking out the open middle third: $[0,1] \to [0,\frac{1}{3}]\cup[\frac{2}{3},1] \to [0,\frac{1}{9}]\cup[\frac{2}{9},\frac{1}{3}]\cup[\frac{2}{3},\frac{7}{9}]\cup[\frac{8}{9},1] \to \ldots$. An explicit formula for the set is: $C = [0,1]\setminus\bigcup_{m=1}^{\infty}\bigcup_{k=0}^{(3^{m-1}-1)} \big]\frac{3k+1}{3^m},\frac{3k+2}{3^m}\big[$.

1.8 Limits and integrals

A few technicalities concerning the definition of limits and the methods of integration will facilitate the discussion of continuous random variables and their distributions. Limits of sequences are required for problems of convergence and for approximating random variables. Taking limits of stochastic variables often needs some care and problems might arise when there are ambiguities in the meaning of limits, which are removed by precise definitions. We introduced already functions like the probability mass function (pmf) and the cumulative probability distribution function (cdf) of discrete random variables, which contain peaks and steps that cannot be subjected to conventional Riemannian integration.

1.8.1 Limits of series of random variables

A sequence of random variables, $\mathcal{X}_n$, is defined on a probability space Ω and it is assumed to have the limit

$$\mathcal{X} \;=\; \lim_{n\to\infty} \mathcal{X}_n \;. \tag{1.45}$$

The probability space Ω, we assume now, has elements ω which have a probability density $p\,(\omega)$. Four different definitions of the limit are common in probability theory [157, pp.40,41].

Almost certain limit: The series $\mathcal{X}_n$ converges *almost certainly* to $\mathcal{X}$ if for all ω except a set of probability zero

$$\mathcal{X}(\omega) \;=\; \lim_{n\to\infty} \mathcal{X}_n(\omega) \;. \tag{1.46}$$

is fulfilled and each realization of $\mathcal{X}_n$ converges to $\mathcal{X}$.

Limit in the mean: The limit in the mean or the mean square limit of a series requires that the mean square deviation of $\mathcal{X}_n(\omega)$ from $\mathcal{X}(\omega)$ vanishes in the limit and the condition is

$$\lim_{n\to\infty} \int_\Omega \mathrm{d}\omega\, p\,(\omega)\Big(\mathcal{X}_n(\omega) - \mathcal{X}(\omega)\Big)^2 \;\equiv\; \lim_{n\to\infty} \big\langle (\mathcal{X}_n - \mathcal{X})^2 \big\rangle \;=\; 0 \;. \tag{1.47}$$

The mean square limit is the standard limit in Hilbert space theory and it is commonly used in quantum mechanics.

Stochastic limit: A limit in probability is called the stochastic limit $\mathcal{X}$ if it fulfils the condition

$$\lim_{n\to\infty} P(|\mathcal{X}_n \,-\, \mathcal{X}| > \varepsilon) \;=\; 0 \tag{1.48a}$$

for any $\varepsilon > 0$. The approach to the stochastic limit is sometimes characterized as convergence in probability

$$\lim_{n\to\infty} \langle f(\mathcal{X}_n)\rangle \ \xrightarrow{P}\ \langle f(\mathcal{X})\rangle \ , \tag{1.48b}$$

where the symbol '$\xrightarrow{P}$' stand for *convergence in probability* (see also section 2.4.3).

Limit in distribution: Probability theory uses also a weaker form of convergence than the previous three limits, the limit in distribution, which requires that for any continuous and bounded function $f(x)$ the relation

$$\lim_{n\to\infty} \langle f(\mathcal{X}_n)\rangle \ \xrightarrow{d}\ \langle f(\mathcal{X})\rangle \tag{1.49}$$

holds, where the symbol '$\xrightarrow{d}$' stand for *convergence in distribution*. An example for convergence in distribution in form of the scoring probability of throwing dice is shown in figure 1.22. This limit, for example, is particularly useful for characteristic functions (section 2.2.3), $\phi(s) = \int_{-\infty}^{\infty} \exp(\dot{\imath}xs) f(x)\,\mathrm{d}x$: If two characteristic functions approach each other, the probability density of $\mathcal{X}_n$ converges to that of $\mathcal{X}$.

Finally we mention stringent conditions for the convergence of functions that are important for probability distributions too. We distinguish *pointwise convergence* and *uniform convergence*. A series of functions $f_0(x), f_1(x), f_2(x), \ldots$ is defined on some interval $I \in \mathbb{R}$. The series converges pointwise to the function $f(x)$ if the limit is fulfilled for every point x:

$$\lim_{n\to\infty} f_n(x) \ = \ f(x) \ \ \forall \ \ x \in I \ . \tag{1.50}$$

It is readily verified that a series of functions can be written as a sum of functions whose convergence is to be tested:

$$\begin{aligned} f(x) \ &= \ \lim_{n\to\infty} f_n(x) \ = \ \lim_{n\to\infty} \sum_{i=1}^{n} g_i(x) \ , \\ g_i(x) \ &= \ \varphi_{i-1}(x) \ - \ \varphi_i(x) \, , \quad \text{and hence} \quad f_n(x) \ = \ \varphi_0(x) \ - \ \varphi_n(x) \ , \end{aligned} \tag{1.51}$$

because $\sum_{i=1}^{n} g_i(x)$ expressed in the functions φ_i is a *telescopic sum*. An example of a series of curves with $\varphi_n(x) = (1 + nx^2)^{-1}$ and accordingly $f_n(x) = nx^2 \,/\, (1+nx^2)$ showing pointwise convergence is shown in figure 1.16. It is easily verified that the limit takes on the form:

$$f(x) \ = \ \lim_{n\to\infty} \frac{nx^2}{1 \ + \ nx^2} \ = \ \begin{cases} 1 & \text{for } x \neq 0 \\ 0 & \text{for } x = 0 \end{cases} \ .$$

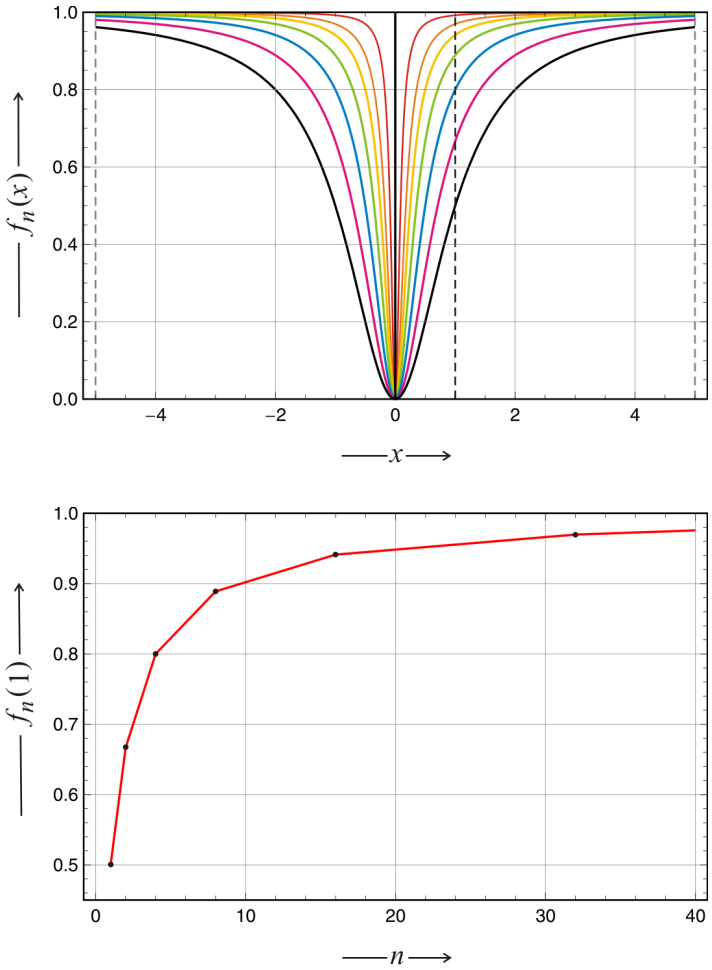


Fig. 1.16 Pointwise convergence. The upper part shows the convergence of the series of functions $f_n(x) = nx^2/(1+nx^2)$ to the limit $\lim_{n\to\infty} f_n(x) = f(x)$ on the real axis $I =]-\infty, \infty[$. The lower plot illustrates the convergence as a function of n at the point $x = 1$. Color code of the upper plot: n=1, black; n=2, violet; n=4, blue; n=8, chartreuse; n=16, yellow; n=32, orange; and n=128, red.

All functions $f_n(x)$ are continuous on the interval $]-\infty, \infty[$ but the limit $f(x)$ is discontinuous at $x = 0$. An interesting historical detail is mentioned here: In 1821 the famous mathematician Augustin Louis Cauchy gave the wrong answer to the question whether or not infinite sums of continuous functions are necessarily continuous and his obvious error had been corrected only thirty years later. It is not hard to imagine that pointwise convergence is compatible with discontinuities in the convergence limit (figure 1.16): At

two neighboring points the convergent series may have very different limits. There are many examples of series of functions, which have a discontinuous infinite limit, two further cases that we shall need later on are $f_n(x) = x^n$ with $I = [0,1] \in \mathbb{R}$ and $f_n(x) = \cos(\pi x)^{2n}$ on $I =]-\infty, \infty[\in \mathbb{R}$.

Uniform convergence is the stronger condition, which guarantees among other things that the limit of a series of continuous functions is *continuous*. It can be defined in terms of equation (1.51): The sum $f_n(x) \sum_{i=1}^{n} g_i(x)$ with $\lim_{n\to\infty} f_n(x) = f(x)$ and $x \in I$ is uniformly convergent in the interval $x \in I$ for every given positive error bound ϵ if there exists a value $\nu \in \mathbb{N}$ such that for any $\nu \geq n$ the relation $|f(x) - f_\nu(x)| < \epsilon$ is fulfilled for all $x \in I$. In compact form the convergence condition may be expressed by

$$\lim_{n\to\infty} \sup\{|f_n(x) - f(x)|\} = 0 \quad \forall \quad x \in I \; . \tag{1.52}$$

A simple but illustrative example is given by the power series on the unit interval, $f(x) = \lim_{n\to\infty} x^n$ with $x \in [0,1]$ which converges pointwise to the discontinuous function $f(x) = 1$ for $x = 1$ and 0 otherwise. A slight modification to $f(x) = \lim_{n\to\infty} x^n/n$ leads to a uniformly converging series, because $f(x) = 0$ is now valid for the entire domain $[0,1]$ (including the point $x = 1$).

1.8.2 Stieltjes integration

Some generalizations of the conventional Riemann integral, which are important in probability theory, are briefly introduced here. In figure 1.17 a sketch is presented that compares Riemann's and the Lebesgue's approach to integration. Stieltjes integration is a generalization of Riemann or Lebesgue integration, which allows for the calculation of integrals over step functions as they occur, for example, in the context of properties derived from cumulative probability distributions. The *Stieltjes integral* is commonly written in the form

$$\int_a^b g(x)\, \mathrm{d}h(x) \; . \tag{1.53}$$

Herein $g(x)$ is the integrand, $h(x)$ is the integrator, and the conventional Riemann integral is retained for $h(x) = x$. The integrator can be visualized best as a weighting function for the integrand. In case $g(x)$ and $h(x)$ are continuous and continuously differentiable the Stieltjes integral can be resolved by partial integration:

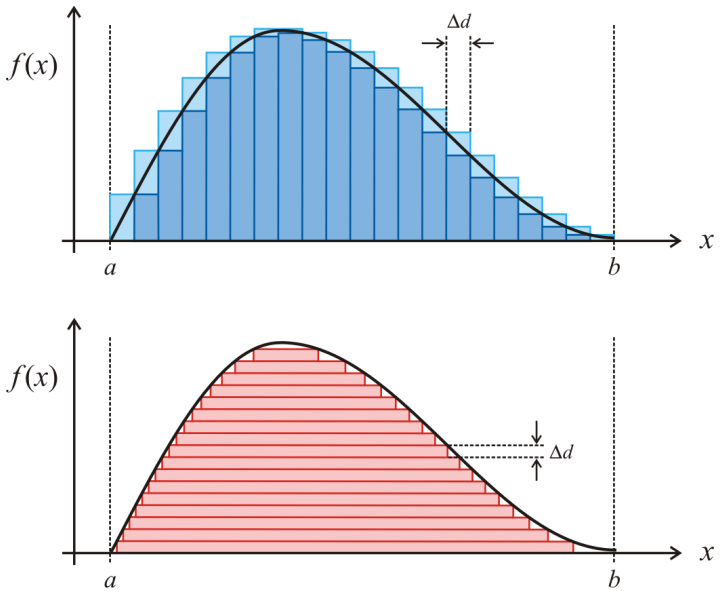


Fig. 1.17 Comparison of Riemann and Lebesgue integrals. In the conventional Riemannian-Darboux integration† the integrand is embedded between an upper sum (light blue) and a lower sum (dark blue) of rectangles. The integral exists iff the upper sum and the lower sum converge to the integrand in the limit $\Delta d \to 0$. The Lebesgue integral can be visualized as an approach to calculating the area enclosed by the x-axis and the integrand through partitioning it into horizontal stripes (red) and considering the limit $\Delta d \to 0$. The definite integral $\int_a^b f(x)\,\mathrm{d}x$ is confining the integrand to a closed interval: $[a, b]$ or $a \leq x \leq b$.

† The concept of representing the integral by the convergence of two sums is due to the French mathematician Gaston Darboux. A function is Darboux integrable iff it is Riemann integrable, and the values of the Riemann and the Darboux integral are equal in case they exist.

$$\begin{aligned}\int_a^b g(x)\,\mathrm{d}h(x) &= \int_a^b g(x)\,\frac{\mathrm{d}h(x)}{\mathrm{d}x}\,\mathrm{d}x \;= \\ &= \Big(g(x)h(x)\Big)\Big|_{x=a}^{b} - \int_a^b \frac{\mathrm{d}g(x)}{\mathrm{d}x}\,h(x)\,\mathrm{d}x \;= \\ &= g(b)h(b) - g(a)h(a) - \int_a^b \frac{\mathrm{d}g(x)}{\mathrm{d}x}\,h(x)\,\mathrm{d}x\;.\end{aligned}$$

The integrator $h(x)$, however, need not be continuous, it may well be a step function $F(x)$, for example a cumulative probability distribution. For $g(x)$ being continuous and $F(x)$ making jumps at the points $x_1, \ldots, x_n \in]a, b[$ with the heights $\Delta F_1, \ldots, \Delta F_n \in \mathbb{R}$, and $\sum_{i=1}^{n} \Delta F_n \leq 1$, respectively, the

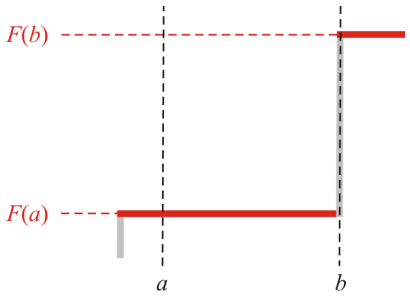


Fig. 1.18 Stieltjes integration of step functions. The figures sketches a Stieltjes integral of a step functions according to the definition of right-hand continuity applied in probability theory (figure 1.10): $\int_a^b \mathrm{d}F(x) = F(b) - F(a) = \Delta F\big|_{x=b}$. The figure also visualizes the Lebesgue-Stieltjes measure $\lambda_F\big((a,b]\big) = F(b) - F(a)$ in (1.60).

Stieltjes integral is of the form

$$\int_a^b g(x)\,\mathrm{d}F(x) \;=\; \sum_{i=1}^{n} g(x_i)\,\Delta F_i \;, \tag{1.54}$$

where the limitation of $\sum_i \Delta F_i$ refers to the normalization of probabilities. With $g(x) = 1$, $b = x$ and in the limit $\lim_{a\to-\infty}$ the integral becomes identical with the (discrete) cumulative probability distribution function (cdf). In figure 1.18 we illustrate the influence of the definition of continuity in probability theory (figure 1.10) on the Stieltjes integral.

Riemann-Stieltjes integration is used in probability theory for the computation of functions of random variables, for example, for the computation of moments of probability densities (section 2.1). If $F(x)$ is the cumulative probability distribution of a random variable $\mathcal{X}$ for the discrete case, the expected value (see section 2.1) for any function $g(\mathcal{X})$ is obtained from

$$E\big(g(\mathcal{X})\big) \;=\; \int_{-\infty}^{\infty} g(x)\,\mathrm{d}F(x) = \;\sum_i g(x_i)\,\Delta F_i \;.$$

If the random variable $\mathcal{X}$ has a probability density $f(x) = \mathrm{d}F(x)/\mathrm{d}x$ with respect to the Lebesgue measure, continuous integration can be used

$$E\big(g(\mathcal{X})\big) \;=\; \int_{-\infty}^{\infty} g(x)\,f(x)\,\mathrm{d}x \;.$$

Important special cases are the moments: $E(\mathcal{X}^n) = \int_{-\infty}^{\infty} x^n\,\mathrm{d}F(x)$.

1.8.3 Lebesgue integration

Lebesgue integration differs from the conventional integration in two aspects: The basis are set theory and measure theory and the integrand is partitioned in horizontal segments whereas Riemannian integration makes use of vertical slices. An important difference for nonnegative functions – like probability functions – between the two integration methods can be visualized in three dimensional space: The volume below a surface given by the function $f(x, y)$ is measured by summation of the volumes of cuboids with squares of edge length Δd, whereas the Lebesgue integral is summing the volumes of layers with thickness Δd between constant level sets. Every continuous bounded function on a compact finite interval, $f \in C[a, b]$, is Riemann integrable and also Lebesgue integrable, and the Riemann and the Lebesgue integrals coincide. The Lebesgue integral is a generalization of the Riemann integral in the sense that certain functions may be Lebesgue integrable in cases where the Riemann integral does not exist. The opposite situations might occur with improper Riemann integrals:[45] Partial sums with alternate signs may converge for the improper Riemann integral whereas Lebesgue integration leads to divergence as shown in case of the alternate harmonic series. The Lebesgue integral can be generalized by the Stieltjes integration technique very much in the same way as the Riemann integral does.

Lebesgue theory of integration assumes the existence of a probability space defined by the triple (Ω, Ξ, μ), which represents the sample space Ω, a σ-algebra Ξ of subsets $A \in \Omega$, and a probability measure $\mu \geq 0$ satisfying $\mu(\Omega) = 1$, respectively. The construction of the Lebesgue integral is similar to the construction of the Riemann integral: The shrinking rectangles (or cuboids in higher dimensions) of Riemannian integration is replaced by horizontal stripes of shrinking height that can be represented by simple functions. Lebesgue integrals on A over nonnegative functions,

$$\int_{\Omega} f \, \mathrm{d}\mu \quad \text{with} \quad f : (\Omega, \Xi, \mu) \to (\mathbb{R}_{\geq 0}, \mathcal{B}, \lambda) \ , \tag{1.55}$$

are defined for measurable functions f, which fulfill

$$f^{-1}\big([a, b]\big) \in \Omega \quad \text{for all} \quad a < b \ . \tag{1.56}$$

This condition is equivalent to the requirement that the pre-image of any Borel subset $[a, b]$ of $\mathbb{R}$ is an element of the event system $\mathcal{B}$. The set of mea-

[45] An improper integral is the limit of a definite integral in a series in which the endpoint of the interval of integration approaches either a finite number b at which the integrand diverges or $\pm\infty$:

$$\int_a^b f(x) \, \mathrm{d}x \ = \ \lim_{\varepsilon \to +0} \int_a^{b-\varepsilon} f(x) \, \mathrm{d}x \quad \text{or} \quad \lim_{b \to \infty} \int_a^b f(x) \, \mathrm{d}x \quad \text{and} \quad \lim_{a \to -\infty} \int_a^b f(x) \, \mathrm{d}x \ .$$

surable functions is closed under algebraic operation and also closed under certain pointwise sequential limits like

$$\sup_{k\in\mathbb{N}} f_k\,, \quad \liminf_{k\in\mathbb{N}} f_k \quad \text{or} \quad \limsup_{k\in\mathbb{N}} f_k\,,$$

which are measurable if the sequence of functions $(f_k)_{k\in\mathbb{N}}$ contains only measurable functions.

The construction of an integral $\int_\Omega f\,\mathrm{d}\mu = \int_\Omega f(x)\,\mu(\mathrm{d}x)$ is done in steps and we apply first the indicator function (1.27):

$$\mathbf{1}_A(x) \;=\; \begin{cases} 1 & \text{iff } x \in A \\ 0 & \text{otherwise} \end{cases} \quad , \tag{1.27a'}$$

to define the integral over $A \in \mathcal{B}^n$ by

$$\int_A f(x)\,dx \;\doteq\; \int \mathbf{1}_A(x)\,f(x)\,dx\,.$$

The indicator function $\mathbf{1}_A$ assigns a volume to Lebesgue measurable sets A by setting $f \equiv 1$

$$\int \mathbf{1}_A\,d\mu \;=\; \mu(A)\,,$$

which is the Lebesgue measure $\mu(A) = \lambda(A)$ for a mapping $\lambda\colon \mathcal{B} \to \mathbb{R}$. It is often useful to consider the expectation value and the variance of the indicator function (1.27)

$$\mathrm{E}\big(\mathbf{1}_A(\omega)\big) \;=\; \frac{A}{\Omega} \;=\; P(A) \quad \text{and} \quad \mathrm{var}\big(\mathbf{1}_A(\omega)\big) \;=\; P(A)\big(1-P(A)\big)\,.$$

We shall make use of this property of the indicator function in section 1.9.2.

Next we define *simple functions*, which are understood as finite linear combinations of indicator functions $g = \sum_j \alpha_j\,\mathbf{1}_{A_j}$ and they are measurable if the coefficients α_j are real numbers and the sets A_j are measurable subsets of Ω. For nonnegative coefficients α_j the linearity property of the integral leads to a measure for nonnegative simple functions:

$$\int \left(\sum_j \alpha_j\,\mathbf{1}_{A_j}\right) \mathrm{d}\mu \;=\; \sum_j \alpha_j \int \mathbf{1}_{A_j}\,\mathrm{d}\mu \;=\; \sum_j \alpha_j\,\mu(A_j)\,.$$

Often a simple function can be written in several ways as a linear combination of indicator functions but the value of the integral will necessarily be the same. Sometimes care is needed for the construction of a real-valued simple function $g = \sum_j \alpha_j \mathbf{1}_{A_j}$ in order to avoid undefined expressions of the kind $\infty - \infty$. Choosing $\alpha_i = 0$ implies that $\alpha_i\,\mu(A_i) = 0$ because $0 \cdot \infty = 0$ by convention in measure theory.

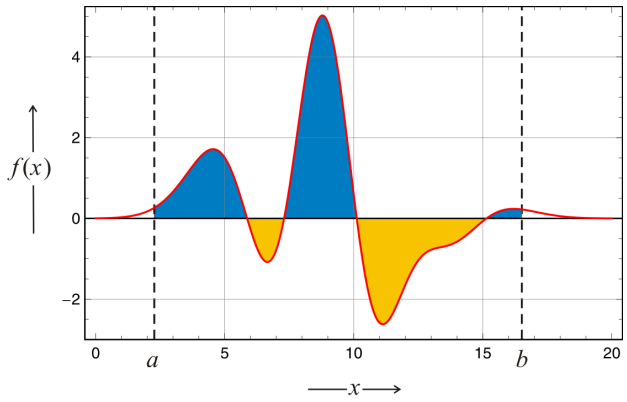


Fig. 1.19 Lebesgue integration of general functions. Lebesgue integration of general functions, i.e. functions with positive and negative stretches, is performed in steps: (i) The integral $I = \int_a^b f \mathrm{d}\mu$ is split into two parts, $I_+ = \int_a^b f_+ \,\mathrm{d}\mu$ (blue) and $I_- = \int_a^b f_- \,\mathrm{d}\mu$ (yellow) function, (ii) the positive part $f_+(x) \doteq \max\{0, f(x)\}$ is Lebesgue integrated like a nonnegative function yielding $I_+ = \int_a^b f_+ \,\mathrm{d}\mu$ and the negative part $f_-(x) \doteq \max\{0, -f(x)\}$ is first mirrored at the x-axis and then Lebesgue integrated like a nonnegative function yielding $I_- = \int_a^b f_- \,\mathrm{d}\mu$, and (iii) the value of the integral is obtained as $I = I_+ - I_-$.

An arbitrary nonnegative function $g : (\Omega, \Xi, \mu) \to (\mathbb{R}_{\geq 0}, \mathcal{B}, \lambda)$ is measurable iff there exists a sequence of simple functions $(g_k)_{k\in\mathbb{N}}$ that converges *pointwise*[46] and growing monotonously to g: $g = \lim_{k\to\infty} g_k$. The Lebesgue integral of a nonnegative and measurable function g is defined by

$$\int_\Omega g \,\mathrm{d}\mu \;=\; \lim_{k\to\infty} \int_\Omega g_k \,\mathrm{d}\mu \tag{1.57}$$

with g_k being simple functions that converge pointwise and monotonously towards g. The limit is independent of the particular choice of the functions g_k. Such a sequence of simple functions is easily visualized, for example, by the bands below the function $g(x)$ in figure 1.17: The band widths Δd decrease and converge to zero as the index increases, $k \to \infty$.

The extension to general functions with positive and negative value domains is straightforward. As shown in figure 1.19 the function to be integrated, $f(x) : [a, b] \to \mathbb{R}$, is split into two regions that many consist of disjoint domains:

[46] Pointwise convergence of a sequence of functions $\{f_n\}$, $\lim_{n\to\infty} f_n = f$ *pointwise* is fulfilled iff $\lim_{n\to\infty} f_n(x) = f(x)$ for every x in the domain (see figure 1.16 and section 1.8.1).

$$f_+(x) \doteq \max\{0, f(x)\}$$
$$f_-(x) \doteq \max\{0, -f(x)\} ,$$

which are considered separately. The function is Lebesgue integrable on the entire domain $[a, b]$ iff both $f_+(x)$ and $f_-(x)$ are Lebesgue integrable and then we have

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^b f_+(x)\,\mathrm{d}x - \int_a^b f_-(x)\,\mathrm{d}x , \tag{1.58}$$

and this yields precisely the same result as obtained for the Riemann integral. Lebesgue integration readily yields the value for the integral of the absolute value of the function

$$\int_a^b |f(x)|\,\mathrm{d}x = \int_a^b f_+(x)\,\mathrm{d}x + \int_a^b f_-(x)\,\mathrm{d}x . \tag{1.59}$$

Whenever the Riemann integral exists it is identical with the Lebesgue integral and for practical purposes the calculation by the conventional technique of Riemannian integration is to be preferred since much more experience is available.

For the purpose of illustration we consider cases where Riemann and Lebesgue integration yield different results. For $\Omega = \mathbb{R}$ and the Lebesgue measure λ holds that functions, which are Riemann integrable on a compact and finite interval $[a, b]$, are Lebesgue integrable too and the values of both integrals are the same, but the inverse is not true: Not every Lebesgue integrable function is Riemann integrable. As an example we consider the Dirichlet step function, $D(x)$, which is the characteristic function of the rational numbers and assumes the value 1 for rational x and the value 0 for irrational x:[47]

$$D(x) = \begin{cases} 1\,, \text{ if } x \in \mathbb{Q}\,, \\ 0\,, \text{ otherwise}\,, \end{cases} \qquad \text{or} \quad D(x) = \lim_{k\to\infty}\Big(\lim_{n\to\infty} \cos^{2n}(k!\,\pi\,x)\Big) .$$

$D(x)$ has no Riemann but a Lebesgue integral. The proof is straightforward: $D(x)$ is lacking Riemann integrability for every arbitrarily small interval: Each partitioning S of the integration domain $[a, b]$ into intervals $[x_{k-1}, x_k]$ leads to parts that contain necessarily at least one rational and one irrational number. Hence the lower Darboux sum,

$$\Sigma_{\mathrm{low}}(S) = \sum_{k=1}^{n} (x_k - x_{k-1}) \cdot \inf_{x_{k-1}<x<x_k} D(x) = 0\,,$$

[47] It is worth noticing that the highly irregular, nowhere continuous Dirichlet function $D(x)$ can be formulated as the (double) pointwise convergence limit of a trigonometric function.

vanishes because the infimum is always zero, and the upper Darboux sum,

$$\Sigma_{\text{high}}(S) = \sum_{k=1}^{n}(x_k - x_{k-1}) \cdot \sup_{x_{k-1}<x<x_k} D(x) = b - a\,,$$

is the length of the integration interval, $b - a = \sum_k (x_k - x_{k-1})$, because the supremum is always one and the summation runs over all partial intervals. Since Riemann integrability requires

$$\sup_S \Sigma_{\text{low}}(S) \;=\; \int_a^b f(x)dx \;=\; \inf_S \Sigma_{\text{high}}(S)$$

$D(x)$ cannot be Riemann integrated.
$D(x)$, on the other hand, has a Lebesgue integral for every interval: $D(x)$ is a nonnegative simple function and therefore we can write the Lebesgue integral over an interval S through sorting into irrational and rational numbers:

$$\int_S D\,\mathrm{d}\lambda \;=\; 0 \cdot \lambda(S \cap \mathbb{R}\backslash\mathbb{Q}) \;+\; 1 \cdot \lambda(S \cap \mathbb{Q})\;,$$

with λ being the Lebesgue measure. The evaluation of the integral is straightforward. The first term vanishes since multiplication by zero yields zero no matter how large $\lambda(S \cap \mathbb{R}\backslash\mathbb{Q})$ is – we recall that $0 \cdot \infty$ is zero by the convention of measure theory – and the second term is also zero as $\lambda(S \cap \mathbb{Q})$ is zero since the set of rational numbers, $\mathbb{Q}$, is countable. Hence we have $\int_S D\,\mathrm{d}\lambda = 0$. □

Another difference between Riemann and Lebesgue integration, however, can occur when the integration is extended to infinity in the improper Riemann integral. Then, the positive and negative contributions may cancel locally in the Riemann summation, whereas divergence may occur in both $f_+(x)$ and in $f_(x)$ since all positive parts and all negative parts are added first in the Lebesgue integral. An example is the improper Riemann integral, $\int_0^\infty \cos x\,\mathrm{d}x$, which has a limit inferior, $\liminf_{n\to\infty} x_n = -1$, and a limit superior, $\limsup_{n\to\infty} x_n = +1$, whereas the corresponding Lebesgue integral does not exist.

A typical example of a function that has an improper Riemann integral but is not Lebesgue integrable is the step function with alternatingly positive and negative stretches of size $\frac{1}{n}$, $(1, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{4}, \ldots)$ (see figure 1.20):
The function $h(x) = (-1)^{k+1}/k$ with $(k-1) \le x < k$ and $k \in \mathbb{N}$ on Riemann integration yields a series of contributions of alternating sign that has a finite infinite sum

$$\int_0^\infty h(x)\,\mathrm{d}x \;=\; 1 - \frac{1}{2} + \frac{1}{3} - \ldots \;=\; \ln 2\;,$$

whereas Lebesgue integrability of h requires $\int_{\mathbb{R}_{\ge 0}} |h|\,\mathrm{d}\lambda < \infty$ and this is not fulfilled since both f_+ and f_- diverge as the harmonic series, $\sum_{k=1}^\infty k^{-1}$, does.

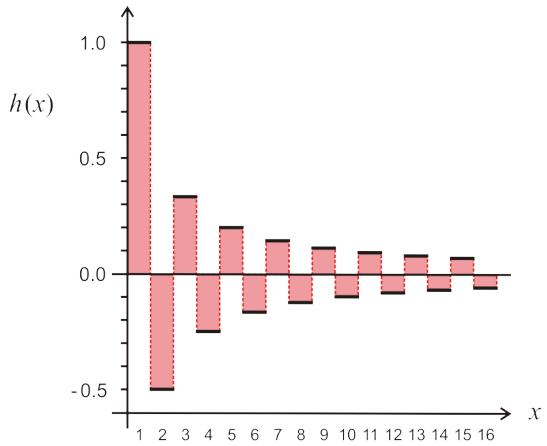


Fig. 1.20 The alternating harmonic series. The alternating harmonic step function, $h(x) = n_k = (-1)^{k+1}/k$ with $(k-1) \leq x < k$ and $n_k \in \mathbb{N}$, has an improper Riemann integral since $\sum_{k=1}^{\infty} n_k = \ln 2$. It is not Lebesgue integrable because the series $\sum_{k=1}^{\infty} |n_k|$ diverges.

The proof is straightforward if one uses Leonhard Euler's result that the series of reciprocal prime number diverges:

$$\sum_{p \text{ prime}} \frac{1}{p} = \frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \frac{1}{13} + \ldots = \infty \,,$$

$$\sum_{o \text{ odd}} \frac{1}{o} = 1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{9} + \frac{1}{11} + \frac{1}{13} + \ldots > \sum_{p \text{ prime}} \frac{1}{p} \,,$$

$$1 + \sum_{e \text{ even}} \frac{1}{e} = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \frac{1}{8} + \frac{1}{10} + \frac{1}{12} + \ldots > \sum_{o \text{ odd}} \frac{1}{o} \,.$$

Since $\infty - 1 = \infty$ both partial sums $\sum_{o \text{ odd}} \frac{1}{o}$ and $\sum_{e \text{ even}} \frac{1}{e}$ and diverge. □

The first case discussed here – no Riemann integral but Lebesgue integrability – is the more important issue since it provides a proof that the set of rational numbers, $\mathbb{Q}$ is of Lebesgue measure zero.

Finally, we introduce the Lebesgue-Stieltjes integral in a way that allows for summarizing the most important results of this section. For each right-hand continuous and monotonously increasing function $F : \mathbb{R} \to \mathbb{R}$ exists a uniquely determined Lebesgue-Stieltjes measure λ_F that fulfils (figure 1.18)

$$\lambda_F\big((a,b]\big) = F(b) - F(a) \quad \text{for all} \quad (a,b] \subset \mathbb{R} \,. \tag{1.60}$$

Such functions $F : \mathbb{R} \to \mathbb{R}$ – being righthand continuous and monotonously increasing – are therefore called *measure generating*. The Lebesgue integral

of a λ_F integrable function f is called Lebesgue-Stieltjes integral

$$\int_A f \, \mathrm{d}\lambda_F \quad \text{with} \quad A \in \mathcal{B} \tag{1.61}$$

being Borel measurable. Let F be the identity function on $\mathbb{R}$,[48]

$$F = \mathrm{id} : \mathbb{R} \to \mathbb{R}, \, \mathrm{id}(x) = x \ ,$$

then the corresponding Lebesgue-Stieltjes measure is the Lebesgue measure itself: $\lambda_F = \lambda_{\mathrm{id}} = \lambda$. For proper Riemann integrable functions f we have stated that the Lebesgue integral is identical with the Riemann integral:

$$\int\limits_{[a,b]} f \, \mathrm{d}\lambda \ = \ \int_a^b f(x) \, \mathrm{d}x \ .$$

The interval $[a, b] = a \leq x \leq b$ is partitioned into a sequence

$$\sigma_n = (a = x_0^{(n)}, x_1^{(n)}, \ldots, x_r^{(n)} = b)$$

where the superscript $'^{(n)}$' indicates a Riemann sum with $|\sigma_n| \to 0$ and the Riemann integral on the righthand side is replaced by the limit of the Riemann summation:

$$\begin{aligned} \int\limits_{[a,b]} f \, \mathrm{d}\lambda \ &= \ \lim_{n\to\infty} \sum_{k=1}^{r} f(x_{k-1}^{(n)}) \left(x_k^{(n)} - x_{k-1}^{(n)} \right) \ = \\ &= \ \lim_{n\to\infty} \sum_{k=1}^{r} f(x_{k-1}^{(n)}) \left(\mathrm{id}(x_k^{(n)}) - \mathrm{id}(x_{k-1}^{(n)}) \right) \ . \end{aligned}$$

The Lebesgue measure λ has been introduced above as the special case $F = \mathrm{id}$ and therefore the Stieltjes-Lebesgue integral is obtained by replacing λ by λ_F and 'id' by F

$$\int\limits_{[a,b]} f \, \mathrm{d}\lambda_F \ = \ \lim_{n\to\infty} \sum_{k=1}^{r} f(x_{k-1}^{(n)}) \left(F(x_k^{(n)}) - F(x_{k-1}^{(n)}) \right) \ .$$

The details of the derivation are found in [64, 322].

In summary, we define a Stieltjes-Lebesgue integral by $(F, f) \colon \mathbb{R} \to \mathbb{R}$, where the two functions F and f are partitioned on the interval $[a, b]$ by the sequence $\sigma = (a = x_0, x_1, \ldots, x_r = b)$:

[48] The identity function $\mathrm{id}(x) \doteq x$ maps a domain, for example $[a, b]$, point by point onto itself.

$$\sum_{\sigma} f \, \mathrm{d}F \;\doteq\; \sum_{k=1}^{r} f(x_{k-1})\big(F(x_k) - F(x_{k-1})\big) \;.$$

The function f is F-integrable on [a,b] if

$$\int_a^b f \, \mathrm{d}F \;=\; \lim_{|\sigma|\to 0} \sum_{\sigma} f \, \mathrm{d}F \tag{1.62}$$

exists in $\mathbb{R}$ and then $\int_a^b f \, \mathrm{d}F$ is called the Stieltjes-Lebesgue integral or sometimes also F-integral of f. In the theory of stochastic processes the Stieltjes-Lebesgue integral is required for the formulation of the Itō integral, which is used in Itō calculus applied to the integration of stochastic differential equations (SDEs; section 3.4) [226, 227].

1.9 Continuous random variables and distributions

Random variables on uncountable sets are completely characterized by a *probability triple* (Ω, Ξ, P). The *triple* is essentially the same as in the case of discrete variables (section 1.6.3) except that the power set $\Pi(\Omega)$ has been replaced by the event system $\Xi \subset \Pi(\Omega)$. We recall that the powerset $\Pi(\Omega)$ is too large for defining probabilities since it contains uncountably many subsets or events A (figure 1.15). The sets in Ξ are the Borel σ-algebras, they are measurable, and they *alone* have probabilities. Accordingly, we are now in the position to handle also probabilities on uncountable sets:

$$\{\omega|\mathcal{X}(\omega) \leq x\} \ \in \ \Xi \ \text{ and } \ P(\mathcal{X} \leq x) \ = \ \frac{|\{\mathcal{X}(\omega) \leq x\}|}{|\Omega|} \tag{1.63a}$$

$$\{a < \mathcal{X} \leq b\} \ = \ \{\mathcal{X} \leq b\} \ - \ \{\mathcal{X} \leq a\} \ \in \Xi \ \text{ with } \ a < b \tag{1.63b}$$

$$P(a < \mathcal{X} \leq b) \ = \ \frac{|\{a < \mathcal{X} \leq b\}|}{|\Omega|} \ = \ F_{\mathcal{X}}(b) \ - \ F_{\mathcal{X}}(a) \ . \tag{1.63c}$$

Equation (1.63a) contains the definition of a real-valued function $\mathcal{X}$ that is called a random variable iff it fulfils $P(\mathcal{X} \leq x)$ for any real number x, equation (1.63b) is valid since Ξ is closed under difference, and finally equation (1.63c) provides the basis for defining and handling probabilities on uncountable sets. The three equations (1.63) together constitute the basis of the probability concept on uncountable sample spaces that will be applied throughout this book.

1.9.1 Densities and distributions

Random variables on uncountable sets Ω are commonly characterized by *probability density functions* (pdf). The probability density function – or density for short – is the continuous analogue to the probability mass function (pmf). A density is a function f on $\mathbb{R} =]-\infty, +\infty[$, $u \to f(u)$, which satisfies the two conditions:[49]

$$\begin{aligned} &\text{(i)} \quad \forall u: \quad f(u) \geq 0 \ , \quad \text{and} \\ &\text{(ii)} \quad \int_{-\infty}^{\infty} f(u)\,\mathrm{d}u \ = \ 1 \ . \end{aligned} \tag{1.64}$$

[49] From here on we shall omit the random variable as subscript and simply write $f(x)$ or $F(x)$ unless a nontrivial specification is required.

Now we can define a class of continuous random variables[50] on general sample spaces: $\mathcal{X}$ is a function on $\Omega : \omega \to \mathcal{X}(\omega)$ whose probabilities are prescribed by means of a density function $f(u)$. For any interval $[a, b]$ the probability is given by

$$P(a \le \mathcal{X} \le b) \;=\; \int_a^b f(u)\,\mathrm{d}u \;. \tag{1.65}$$

If A is the union of not necessarily disjoint intervals – some of which may be even infinite – the probability can be derived in general from the density

$$P(\mathcal{X} \in A) \;=\; \int_A f(u)\,\mathrm{d}u \;,$$

in particular, A can be split in disjoint intervals, $A = \bigcup_{j=1}^{k} [a_j, b_j]$ and then the integral can be rewritten as

$$\int_A f(u)\,\mathrm{d}u \;=\; \sum_{j=1}^{k} \int_{a_j}^{b_j} f(u)\,\mathrm{d}u \;.$$

For the interval $A =]-\infty, x]$ we define the *cumulative probability distribution function* (cdf) $F(x)$ of the continuous random variable $\mathcal{X}$

$$F(x) \;=\; P(\mathcal{X} \le x) \;=\; \int_{-\infty}^{x} f(u)\,\mathrm{d}u \;.$$

An easy to verify and useful relation defines the complementary cumulative distribution function (ccdf):

$$\tilde{F}(x) \;=\; P(\mathcal{X} < x) \;=\; 1 \,-\, F(x) \;. \tag{1.66}$$

If f is continuous then it is the derivative of F as follows from the fundamental theorem of calculus

$$F'(x) \;=\; \frac{\mathrm{d}F(x)}{\mathrm{d}x} \;=\; f(x).$$

If the density f is not continuous everywhere, the relation is still true for every x at which f is continuous.

If the random variable $\mathcal{X}$ has a density, then we find by setting $a = b = x$

$$P(\mathcal{X} = x) \;=\; \int_x^x f(u)\,\mathrm{d}u \;=\; 0$$

reflecting the trivial geometric result that every line segment has zero area. It seems somewhat paradoxical that $\mathcal{X}(\omega)$ must be some number for every ω whereas any given number has probability zero. The paradox is resolved

[50] Random variables having a density are often called *continuous* in order to distinguish them from *discrete* random variables defined on countable sample spaces.

by looking at countable and uncountable sets in more depth as we did in sections 1.5 and 1.6.4.

As an illustrative example for continuous probability functions we present here the *normal distribution*, which is of primary importance in probability theory for several reasons: (i) It is mathematically simple and well behaved, (ii) it is exceedingly *smooth* as it can be differentiated an infinite number of times, and (iii) all distributions converge to the normal distribution in the limit of large sample numbers as expressed by the *central limit theorem* (subsection 2.4.2). The density of the normal distribution is a Gaussian function named after the German mathematician Carl Friedrich Gauß and is also called symmetric bell curve.

$$\mathcal{N}(x;\mu,\sigma^2): \quad f(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\;\exp\left(-\frac{(x-\mu)^2}{2\,\sigma^2}\right)\;, \tag{1.67}$$

$$F(x) \;=\; \frac{1}{2}\left(1+\operatorname{erf}\Big(\frac{x-\mu}{\sqrt{2\sigma^2}}\Big)\right)\;. \tag{1.68}$$

Herein 'erf' is the error function.[51] This function and its complement, 'erfc', are defined by

$$\operatorname{erf}(x) \;=\; \frac{2}{\sqrt{\pi}}\int_0^x e^{-z^2}\mathrm{d}z \quad \text{and} \quad \operatorname{erfc}(x) \;=\; \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-z^2}\mathrm{d}z$$

The two parameters of the normal distribution, μ and σ^2, are the expectation value and the variance of a normally distributed random variable, respectively, and σ is called the mean deviation.

Although the central limit theorem will be discussed separately in section 2.4.2, we present here an example for the convergence of a probability distribution towards the normal distribution we are already familiar with: the rolling dice problem extended to n dice. A collection of n dice is thrown simultaneously and the total score of all dice together is recorded (figure 1.22). The probability of a total score of k points obtained through rolling n dice with s faces can be calculated by means of combinatorics:

$$f_{s,n}(k) \;=\; \frac{1}{s^n}\sum_{i=0}^{\lfloor\frac{k-n}{s}\rfloor}(-1)^i\binom{n}{i}\binom{k-s\,i-1}{n-1} \tag{1.69}$$

The results for small values of n and ordinary dice ($s = 6$) are illustrated in figure 1.22. The convergence to a continuous probability density is nicely

[51] We remark that $\operatorname{erf}(x)$ and $\operatorname{erfc}(x)$ are not normalized in the same way as the normal density: $\operatorname{erf}(x)+\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_0^\infty \exp(-t^2)\,\mathrm{dt} = 1$, but $\int_0^\infty f(x)\mathrm{d}x = \frac{1}{2}\int_{-\infty}^\infty f(x)\mathrm{d}x = \frac{1}{2}$.

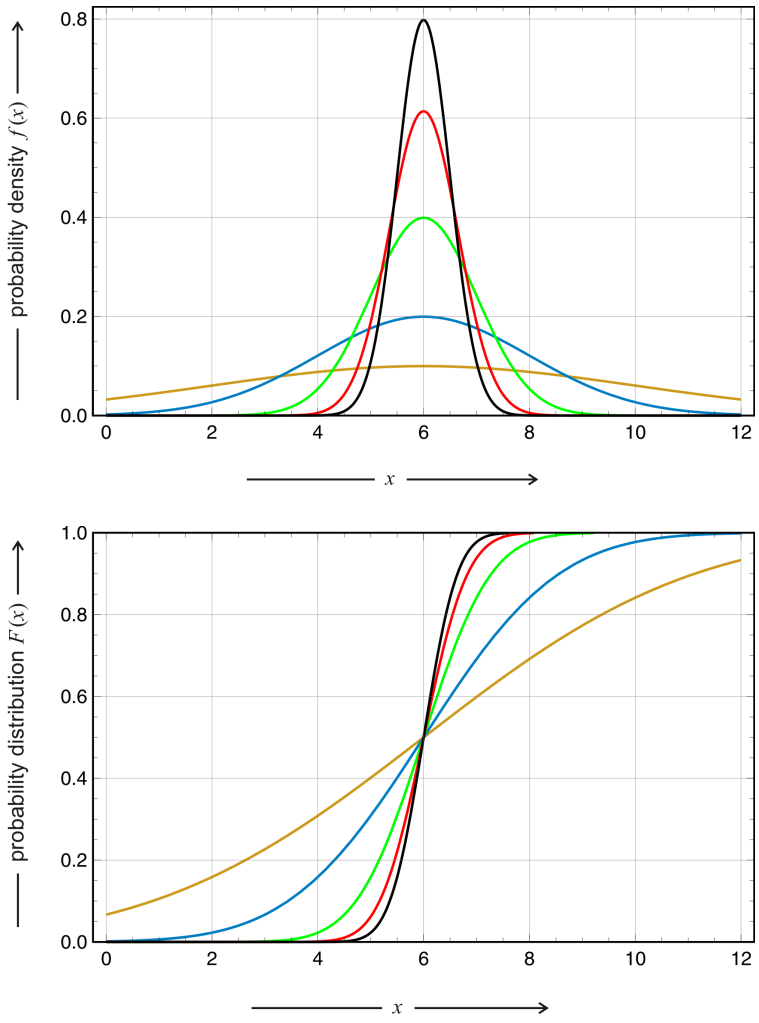


Fig. 1.21 Normal density and distribution. In the plots the normal distribution, $\mathcal{N}(\mu, \sigma)$, is shown in from of the probability density $f(x) = \exp\left(-(x-\mu)^2/(2\sigma^2)\right) / (\sqrt{2\pi}\,\sigma)$ and the probability distribution $F(x) = \left(1 + \mathrm{erf}\left((x-\mu)/\sqrt{2\sigma^2}\right) / 2\right)$ where 'erf' represents the error function. Choice of parameters: $\mu = 6$ and $\sigma = 0.5$ (black), 0.65 (red), 1 (green), 2 (blue) and 4 (yellow).

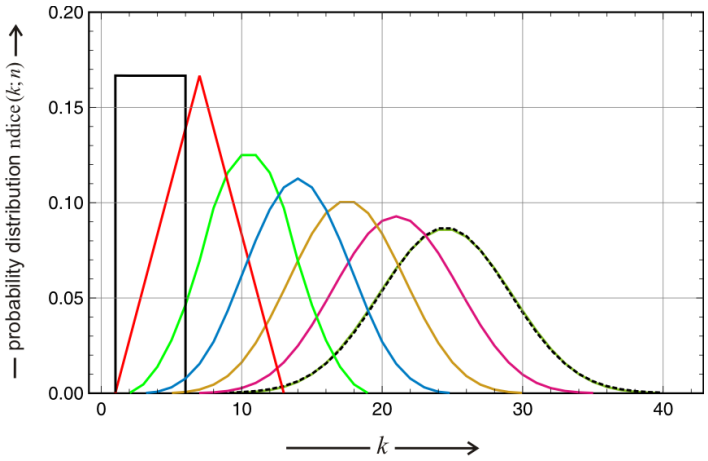


Fig. 1.22 Convergence of the probability mass function for rolling n dice to the normal density. The series of probability mass functions for rolling n dice, $f_{nd}(k)$, begins with a pulse function $f_{1d}(k) = 1/6$ for $i = 1, \ldots, 6$ ($n = 1$), next comes a tent function ($n = 2$), and then follows a gradual approach towards the normal distribution, ($n = 3, 4, \ldots$). For $n = 7$ we show the fitted normal distribution (broken black curve) coinciding almost perfectly with $f_{7d}(k)$. The series of densities provides an example for convergence in distribution (section 1.8.1). Choice of parameters: $s = 6$ and $n = 1$ (black), 2 (red), 3 (green), 4 (blue), 5 (yellow), 6 (magenta), and 7 (chartreuse).

illustrated. For $n = 7$ the deviation from a the Gaussian curve of the normal distribution is hardly recognizable.

Sometimes it is useful to discretize a density function in order to yield a set of elementary probabilities. The x-axis is divided up into m pieces (figure 1.23), not necessarily equal and not necessarily small, and we denote the piece of the integral on the interval $\Delta_k = x_{k+1} - x_k$, i.e. between the values $u(x_k)$ and $u(x_{k+1})$ of the variable u, by

$$p_k \;=\; \int_{x_k}^{x_{k+1}} f(u)\,\mathrm{d}u\,,\;\; 0 \le k \le m-1\;, \tag{1.70}$$

where the p_k-values fulfil.

$$\forall\; k:\;\; p_k \ge 0 \;\;\text{ and }\;\; \sum_{k=0}^{m-1} p_k \;=\; 1\;.$$

If we choose $x_0 = -\infty$ and $x_m = +\infty$ we are dealing with a partition that is not finite but countable, provided we label the intervals suitably, for example $\ldots, p_{-2}, p_{-1}, p_0, p_1, p_2, \ldots$. Now we consider a random variable $\mathcal{Y}$ such that

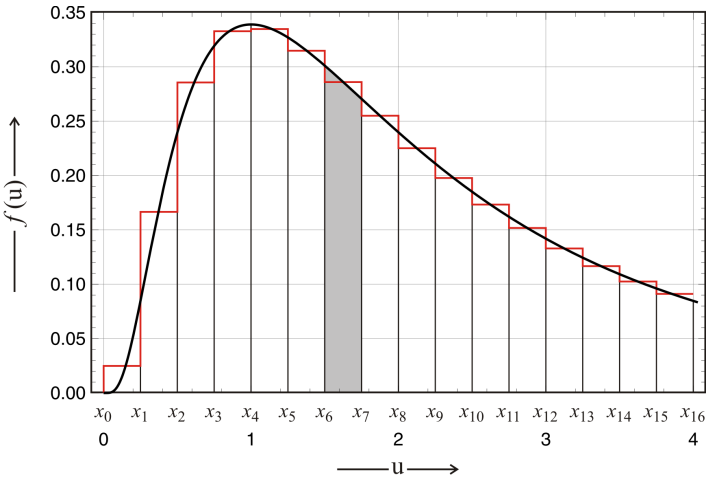


Fig. 1.23 Discretization of a probability density. A segment $[x_0, x_m]$ on the u-axis is divided up into m not necessarily equal intervals and elementary probabilities are obtained by integration. The curve shown here is the density of the lognormal distribution $\ln\mathcal{N}(\nu, \sigma^2)$:

$$f(u) \;=\; \frac{1}{u\,\sqrt{2\pi\,\sigma^2}}\, e^{-(\ln u - \nu)^2/(2\sigma^2)}.$$

The red step function represents the discretized density. The hatched area is the probability $p_6 = \int_{x_6}^{x_7} f(u)\,\mathrm{d}u$ with the parameters $\nu = \ln 2$ and $\sigma = \sqrt{\ln 2}$.

$$P(\mathcal{Y} = x_k) \;=\; p_k \;, \tag{1.70'}$$

where we may replace x_k by any value of x in the subinterval $[x_k, x_{k+1}]$. The random variable $\mathcal{Y}$ can be interpreted as the discrete analogue of the random variable $\mathcal{X}$. Making the intervals Δ_k smaller increases the accuracy of the approximation through discretization and this procedure has a lot in common with Riemann integration.

1.9.2 Expectation values and variances

Although we shall treat expectation values and other moments of probability distributions in the next chapter 2, we make here a short digression in order to present examples for various integration concepts. The calculation of expectation values and variances from continuous densities is straightforward

$$\mathrm{E}(\mathcal{X}) \;=\; \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x \;=\; \int_{0}^{\infty} \big(1-F(x)\big)\,\mathrm{d}x \;-\; \int_{-\infty}^{0} F(x)\,\mathrm{d}x \;\text{ and } \quad (1.71a)$$

$$\mathrm{var}(\mathcal{X}) \;=\; \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x \;-\; \mathrm{E}(\mathcal{X})^2\;. \quad (1.71b)$$

The computation of the expectation value from the probability distribution is the analogue to the discrete case (1.21a). We present the derivation of the expression here as an exercise in handling probabilities and integrals [187]. Like in a Lebesgue integral we decompose $\mathcal{X}$ into positive and negative parts: $\mathcal{X} = \mathcal{X}^+ - \mathcal{X}^-$ with $\mathcal{X}^+ = \max\{\mathcal{X}, 0\}$and $\mathcal{X}^- = \max\{-\mathcal{X}, 0\}$. Then, we express both parts by means of indicator functions

$$\mathcal{X}^+ \;=\; \int_{0}^{\infty} \mathbf{1}_{\mathcal{X}>\vartheta}\,\mathrm{d}\vartheta \quad \text{and} \quad \mathcal{X}^- \;=\; \int_{-\infty}^{0} \mathbf{1}_{\mathcal{X}\le\vartheta}\,\mathrm{d}\vartheta\;.$$

By applying Fubini's theorem named after the Italian mathematician Guido Fubini [153] we reverse the order of taking the expectation value and integration, make use of (1.27b) and (1.66), and find

$$\begin{aligned}
\mathrm{E}(\mathcal{X}) \;&=\; \mathrm{E}(\mathcal{X}^+ - \mathcal{X}^-) \;=\; \mathrm{E}(\mathcal{X}^+) - \mathrm{E}(\mathcal{X}^-) \;= \\
&=\; \mathrm{E}\left(\int_{0}^{\infty} \mathbf{1}_{\mathcal{X}>\vartheta}\,\mathrm{d}\vartheta\right) \;-\; \mathrm{E}\left(\int_{-\infty}^{0} \mathbf{1}_{\mathcal{X}\le\vartheta}\,\mathrm{d}\vartheta\right) \;= \\
&=\; \int_{0}^{\infty} \mathrm{E}\left(\mathbf{1}_{\mathcal{X}>\vartheta}\right)\,\mathrm{d}\vartheta \;-\; \int_{-\infty}^{0} \mathrm{E}\left(\mathbf{1}_{\mathcal{X}\le\vartheta}\right)\,\mathrm{d}\vartheta \;= \\
&=\; \int_{0}^{\infty} P(\mathcal{X}>\vartheta)\,\mathrm{d}\vartheta \;-\; \int_{-\infty}^{0} P(\mathcal{X}\le\vartheta)\,\mathrm{d}\vartheta \;= \\
&=\; \int_{0}^{\infty} \big(1-F(\vartheta)\big)\,\mathrm{d}\vartheta \;-\; \int_{-\infty}^{0} F(\vartheta)\,\mathrm{d}\vartheta\,. \qquad \square
\end{aligned}$$

The calculation of expectation values directly from the cumulative distribution function has the advantage to be applicable also in cases where densities do not exist or where they are hard to handle.

1.9.3 Continuous variables and independence

In the joint distribution function of the random vector $\vec{\mathcal{V}} = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$ independence is tantamount to factorizability:

$$F(x_1, \ldots, x_n) \;=\; F_1(x_1) \cdot \ldots \cdot F_n(x_n)\;,$$

where the F_j's are the marginal distributions of the random variables, the $\mathcal{X}_j$'s ($1 \leq j \leq n$). As in the discrete case the marginal distributions are sufficient to calculate joint distributions of independent random variables.

For the continuous case we can formulate the definition of independence for sets $S_1, \ldots, S_n$ forming a Borel family. In particular, when there is a joint density function $f(u_1, \ldots, u_n)$, we have

$$\begin{aligned} P(\mathcal{X}_1 \in S_1, \ldots, \mathcal{X}_n \in S_n) &= \int_{S_1} \cdots \int_{S_n} f(u_1, \ldots, u_n)\, du_1 \ldots du_n = \\ &= \int_{S_1} \cdots \int_{S_n} f_1(u_1) \ldots f_n(u_n)\, du_1 \ldots du_n = \\ &= \left(\int_{S_1} f_1(u_1)\, du_1 \right) \cdot \ldots \cdot \left(\int_{S_n} f_n(u_n)\, du_n \right) , \end{aligned}$$

where $f_1, \ldots, f_n$ are the marginal densities, for example

$$f_1(u_1) = \int_{S_2} \cdots \int_{S_n} f(u_1, \ldots, u_n)\, du_2 \ldots du_n , \tag{1.72}$$

and eventually we find for the density case:

$$f(u_1, \ldots, u_n) = f_1(u_1) \ldots f_n(u_n) . \tag{1.73}$$

As we have seen here, stochastic independence is the basis for factorization of joint probabilities, distributions, densities, and other functions. Independence is a stronger criterium than uncorrelatedness 2.3.4 as we shall show in section 2.3.4.

1.9.4 Probabilities of discrete and continuous variables

A comparison of the formalisms of probability theory on countable and uncountable sample spaces closes this chapter. For this goal we repeat and compare in table 1.3 the basic features of discrete and continuous probability distributions as they have been discussed in section 1.6.3 and 1.9.1, respectively.

Discrete probability distribution are defined on countable sample spaces and their random variables are discrete sets of events $\omega \in \Omega$, for example sample points on an closed interval $[a, b]$:

$$\{a \leq \mathcal{X} \leq b\} = \{\omega \,|\, a \leq \mathcal{X} \leq b\} .$$

If the sample space Ω is finite or countable infinite, the exact range of $\mathcal{X}$ is a set of real numbers w_i

Table 1.3 Comparison of the formalism of probability theory on countable and uncountable sample spaces. The table shows the basic formulas for discrete and continuous random variables.

	Expression	Countable case		Uncountable case		
Domain, full	$A \in \Omega$	$w_n,\ n = \ldots, -2, -1, 0, 1, 2, \ldots$	$w_n \in \mathbb{Z}$	$-\infty < u < \infty$	$]-\infty, \infty[$	$u \in \mathbb{R}$
nonnegative		$w_n,\ n = 0, 1, 2, \ldots$	$w_n \in \mathbb{N}$	$0 \leq u < \infty$	$[0, \infty[$	$u \in \mathbb{R}_{\geq 0}$
positive		$w_n,\ n = 1, 2, \ldots$	$w_n \in \mathbb{N}$	$0 < u < \infty$	$[0, \infty[$	$u \in \mathbb{R}_{>0}$
Probability	$P(\mathcal{X} \in A) : a \in \Omega$	p_n		$dF(u) = f(u)\,\mathrm{d}u$		
Interval	$P(a \leq \mathcal{X} \leq b)$	$\sum_{a \leq w_n \leq b} p_n$		$\int_a^b f(u)\,\mathrm{d}u$		
Density, pmf or pdf	$f(x) = P(\mathcal{X} = x)$	$\begin{cases} p_n & \text{if } x \in W_\mathcal{X} \\ 0 & \text{if } x \notin W_\mathcal{X} \end{cases}$		$f(u)\,\mathrm{d}u$		
Distribution, cdf	$F(x) = P(\mathcal{X} \leq x)$	$\sum_{w_n \leq x} p_n$		$\int_{-\infty}^x f(u)\,\mathrm{d}u$		
Expectation value	$\mathrm{E}(\mathcal{X}) = \sum_n n f_\mathcal{X}(n)$	$\sum_n p_n w_n$	$\sum_n p_n \lvert w_n \rvert < \infty$	$\int_{-\infty}^\infty u f(u)\,\mathrm{d}u$		$\int_{-\infty}^\infty \lvert u \rvert f(u)\,\mathrm{d}u < \infty$
		$\mathrm{E}(\mathcal{X}) = \sum_n \big(1 - F(n)\big)$	$n \in \mathbb{N}$	$\int_0^\infty \big(1 - F(u)\big)\,\mathrm{d}u$		$u \in \mathbb{R}_{\geq 0}$
Variance	$\mathrm{var}(\mathcal{X}) = \sum_n n^2 f_\mathcal{X}(n) - \mathrm{E}(\mathcal{X})^2$	$\sum_n p_n w_n^2 - \mathrm{E}(\mathcal{X})^2$	$\sum_n p_n\, w_n^2 < \infty$	$\int_{-\infty}^\infty u^2 f(u)\,\mathrm{d}u - \mathrm{E}(\mathcal{X})^2$		$\int_{-\infty}^\infty u^2 f(u)\,\mathrm{d}u < \infty$
		$2\sum_n k\big(1 - F(n)\big) - \mathrm{E}(\mathcal{X})^2$	$n \in \mathbb{N}$	$2\int_0^\infty u\big(1 - F(u)\big)\,\mathrm{d}u - \mathrm{E}(\mathcal{X})^2$		$u \in \mathbb{R}_{\geq 0}$

$$W_{\mathcal{X}} \;=\; \{w_1, w_2, \ldots, w_n, \ldots\} \;\text{ with }\; w_k \in \Omega \,\forall\, k \;.$$

Introducing probabilities for individual events, $p_n = P\big(\mathcal{X} = w_n | \, w_n \in W_{\mathcal{X}}\big)$ and $P\big(\mathcal{X}(x) = 0 | x \notin W_{\mathcal{X}}\big)$, yields

$$P(\mathcal{X} \in A) \;=\; \sum_{w_n \in A} p_n \;\text{ with }\; A \in \Omega$$

or, in particular,

$$P(a \le \mathcal{X} \le b) \;=\; \sum_{a \le w_n \le b} p_n \;. \tag{1.31}$$

Two probability functions are in common use, the probability mass function (pmf)

$$f_{\mathcal{X}}(x) \;=\; P(\mathcal{X} = x) \begin{cases} p_n & \text{if } x = w_n \in W_{\mathcal{X}} \;, \\ 0 & \text{if } x \neq W_{\mathcal{X}} \;, \end{cases}$$

and the cumulative distribution function (cdf)

$$F_{\mathcal{X}}(x) \;=\; P(\mathcal{X} \,\le\, x) \;=\; \sum_{w_n \le x} p_n \;,$$

with two properties following form the property of probabilities:

$$\lim_{x \to -\infty} F_{\mathcal{X}}(x) \;=\; 0 \quad \text{and} \quad \lim_{x \to +\infty} F_{\mathcal{X}}(x) \;=\; 1 \;.$$

Continuous probability distributions are defined on uncountable, Borel measurable sample spaces and their random variables $\mathcal{X}$ have densities. A *probability density function* (pdf) is a mapping

$$f : \mathbb{R} \implies \mathbb{R}_{\ge 0} \;,$$

which satisfies the two conditions:

$$\begin{aligned} &\text{(i)} \qquad f(u) \ge 0 \;\forall\; u \in \mathbb{R} \;, \quad \text{and} \\ &\text{(ii)} \qquad \int_{-\infty}^{\infty} f(u)\, du \;=\; 1 \;. \end{aligned} \tag{1.74}$$

Random variables $\mathcal{X}$ are functions on Ω: $\omega \Longrightarrow \mathcal{X}(\omega)$ whose probabilities are derived from density functions $f(u)$:

$$P(a \le \mathcal{X} \le b) \;=\; \int_a^b f(u)\, du \;. \tag{1.65}$$

As in the discrete case the probability functions come in two forms: (i) the probability density function (pdf) defined above,

$$dF(u) \;=\; f(u)\,du\;,$$

and (ii) the cumulative distribution function (cdf)

$$F(x) \;=\; P(\mathcal{X} \leq x) \;=\int_{-\infty}^{x} f(u)\,\mathrm{d}u \;\text{ with }\; \frac{\mathrm{d}F(x)}{\mathrm{dx}} \;=\; f(x)$$

provided the function $f(x)$ is continuous.

Conventional thinking in terms of probabilities has been extended in two important ways in the last two sections: (i) Handling of uncountable sets allowed for definition of and calculation with probabilities when comparison by counting is not possible and (ii) Lebesgue-Stieltjes integration provided an extension of calculus to the step functions encountered with discrete probabilities.

Chapter 2
Distributions, moments, and statistics

Make things as simple as possible but not simpler.
Albert Einstein 1950.

Abstract . The moments of probability distributions represent the link between theory and observations since they are readily accessible to measurement. Generating functions looking rather abstract became important as highly versatile concepts and tools for solving specific problems. The probability distributions, which are most important in application are reviewed. Then the central limit theorem being the basis of the law of large numbers is presented and the chapter is closed by discussing real world samples that cover a part sometimes only a small of sample space.

In this chapter we make an attempt to bring probability theory closer to applications. Probability distributions and densities are used to calculate measurable quantities like expectation values, variances, and higher moments. The moments provide partial information on the probability distributions since the full information would require a series expansion up to infinite order.

2.1 Expectation values and higher moments

Random variables are accessible to analysis via their probability distributions. Straightforward and full information is derived from ensembles defined on entire sample space Ω. Complete coverage, of course, is an ideal reference that can never be achieved in real situations. Samples obtained in experimental observations are commonly much smaller than an exhaustive collection. We begin here with a discussion of the theoretical reference and introduce mathematical statistics afterwards. Distributions can be characterized by moments that are powers of variables $\mathcal{X}^r$ averaged over sample space. Most important are the first two moments having a straightforward interpretation: The expectation value $\mathrm{E}(\mathcal{X})$ is the average value of a distribution and the variance $\mathrm{var}(\mathcal{X})$ or $\sigma^2(\mathcal{X})$ is a measure of the width of a distribution.

2.1.1 First and second moments

The most natural and important ensemble property is the *expectation value* or *average* written $\mathrm{E}(\mathcal{X})$ or as preferred in physics $\langle\mathcal{X}\rangle$. We begin with a countable sample space Ω:

$$\mathrm{E}(\mathcal{X}) \;=\; \sum_{\omega\in\Omega} \mathcal{X}(\omega)P(\omega) \;=\; \sum_{n} w_n\, p_n \;. \tag{2.1}$$

In the most common special case of a random variable $\mathcal{X}$ on $\mathbb{N}$ we have $w_n = n$ and find

$$\mathrm{E}(\mathcal{X}) \;=\; \sum_{n=0}^{\infty} n\, p_n \;.$$

The expectation value (2.1) exists when the series converges in absolute values, $\sum_{\omega\in\Omega} |\mathcal{X}(\omega)|\, P(\omega) < \infty$. Whenever the random variable $\mathcal{X}$ is bounded, which means that there exists a number m such that $|\mathcal{X}(\omega)| \le m$ for all $\omega \in \Omega$, then it is summable and in fact

$$\mathrm{E}(|\mathcal{X}|) \;=\; \sum_{\omega} |\mathcal{X}(\omega)|\, P(\omega) \;\le\; m \sum_{\omega} P(\omega) \;=\; m \;.$$

It is straightforward to show that the sum of two summable random variables, $\mathcal{X} + \mathcal{Y}$, is summable, and the expectation value of the sum is the sum of the expectation values:

$$\mathrm{E}(\mathcal{X} + \mathcal{Y}) \;=\; \mathrm{E}(\mathcal{X}) \,+\, \mathrm{E}(\mathcal{Y}) \;.$$

The relation can be extended to an arbitrary countable number of random variables:

$$\mathrm{E}\left(\sum_{k=1}^{n} \mathcal{X}_k\right) \;=\; \sum_{k=1}^{n} \mathrm{E}(\mathcal{X}_k) \;.$$

In addition, the expectation values fulfill the following relations $\mathrm{E}(a) = a$, $\mathrm{E}(a\mathcal{X}) = a \cdot \mathrm{E}(\mathcal{X})$ which can be combined in

$$\mathrm{E}\left(\sum_{k=1}^{n} a_k\, \mathcal{X}_k\right) \;=\; \sum_{k=1}^{n} a_k \cdot \mathrm{E}(\mathcal{X}_k) \;. \tag{2.2}$$

Accordingly, $\mathrm{E}(\cdot)$ fulfils the conditions for a *linear operator*.

For a random variable $\mathcal{X}$ on an arbitrary sample space Ω the expectation value may be written as an abstract integral on Ω or as an integral over $\mathbb{R}$ provided the density $f(u)$ exists:

$$\mathrm{E}(\mathcal{X}) \;=\; \int_{\Omega} \mathcal{X}(\omega)\, d\omega \;=\; \int_{-\infty}^{+\infty} u\, f(u)\, \mathrm{d}u \;. \tag{2.3}$$

It is worth to reconsider the discretization of a continuous density (figure 1.23) in this context: The discrete expression for the expectation value is based upon $p_n = P(\mathcal{Y} = x_n)$ as outlined in equations (1.70) and (1.70'),

$$\mathrm{E}(\mathcal{Y}) \;=\; \sum_n x_n\, p_n \;\approx\; \mathrm{E}(\mathcal{X}) \;=\; \int_{-\infty}^{+\infty} u\, F(u)\,\mathrm{d}u \;,$$

and approximates the exact value similarly as the Darboux sum does in case of a Riemann integral.

For two or more variables, for example $\vec{\mathcal{V}} = (\mathcal{X}, \mathcal{Y})$ described by a joint density $f(u, v)$, we have

$$\mathrm{E}(\mathcal{X}) \;=\; \int_{-\infty}^{+\infty} u\, f(u,*)\,\mathrm{d}u \;\;\text{and}\;\; \mathrm{E}(\mathcal{Y}) \;=\; \int_{-\infty}^{+\infty} v\, f(*,v)\,\mathrm{d}v \;,$$

where $f(u,*) = \int_{-\infty}^{+\infty} f(u,v)\,\mathrm{d}v$ and $f(*,v) = \int_{-\infty}^{+\infty} f(u,v)\,\mathrm{d}u$ are the marginal densities.

The expectation value of the sum of the variables, $\mathcal{X}+\mathcal{Y}$, can be evaluated by iterated integration:

$$\begin{aligned}
\mathrm{E}(\mathcal{X}+\mathcal{Y}) \;&=\; \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (u+v)\, f(u,v)\,\mathrm{d}u\,\mathrm{d}v \;= \\
&=\; \int_{-\infty}^{+\infty} u\,\mathrm{d}u \left(\int_{-\infty}^{+\infty} f(u,v)\,\mathrm{d}v\right) + \int_{-\infty}^{+\infty} v\,\mathrm{d}v \left(\int_{-\infty}^{+\infty} f(u,v)\,\mathrm{d}u\right) \;= \\
&=\; \int_{-\infty}^{+\infty} u\, f(u,*)\,\mathrm{d}u \;+\; \int_{-\infty}^{+\infty} v\, f(*,v)\,\mathrm{d}v \;= \\
&=\; \mathrm{E}(\mathcal{X}) \;+\; \mathrm{E}(\mathcal{Y}) \;,
\end{aligned}$$

which establishes the expression previously derived in the discrete case.

The *multiplication theorem* of probability theory requires that the two variables $\mathcal{X}$ and $\mathcal{Y}$ are independent and summable and this implies for the discrete and the continuous case,[1]

$$\mathrm{E}(\mathcal{X}\cdot\mathcal{Y}) \;=\; \mathrm{E}(\mathcal{X})\cdot\mathrm{E}(\mathcal{Y}) \;\;\text{and} \tag{2.4a}$$

$$\begin{aligned}
\mathrm{E}(\mathcal{X}\cdot\mathcal{Y}) \;&=\; \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} u\, v\, f(u,v)\,\mathrm{d}u\,\mathrm{d}v \;= \\
&=\; \int_{-\infty}^{+\infty} u\, f(u,*)\,\mathrm{d}u \int_{-\infty}^{+\infty} v\, f(*,v)\,\mathrm{d}v \;= \\
&=\; \mathrm{E}(\mathcal{X})\cdot\mathrm{E}(\mathcal{Y}) \;,
\end{aligned} \tag{2.4b}$$

[1] A proof is found in [69, pp.164-166].

respectively. The multiplication theorem is easily extended to any finite number of independent and summable random variables:

$$\mathrm{E}(\mathcal{X}_1, \ldots, \mathcal{X}_n) \;=\; \mathrm{E}(\mathcal{X}_1) \cdot \ldots \cdot \mathrm{E}(\mathcal{X}_n) \;. \tag{2.4c}$$

Let us now consider the expectation values of special functions of random variables, in particular their powers $\mathcal{X}^r$, which give rise to the *raw moments* of the probability distribution, $\hat{\mu}_r$. For a random variable $\mathcal{X}$ we distinguish the r-th moments $\mathrm{E}(\mathcal{X}^r)$ and the so-called *centered moments*[2] $\mu_r = \mathrm{E}(\tilde{\mathcal{X}}^r)$ referring to the random variable

$$\tilde{\mathcal{X}} \;=\; \mathcal{X} \,-\, \mathrm{E}(\mathcal{X}) \;.$$

Clearly, the first raw moment is the expectation value and the first centered moment vanishes, $\mathrm{E}(\tilde{\mathcal{X}}) = \mu_1 = 0$. Often the expectation value is denoted by $\mu = \hat{\mu}_1 = \mathrm{E}(\mathcal{X}) = \langle \mathcal{X} \rangle$, a notation that we shall use too for the sake of convenience but it is important not to confuse μ and μ_1.

In general, a moment is defined about some point a by means of the random variable

$$\mathcal{X}^{(a)} \;=\; \mathcal{X} \,-\, a \;.$$

For $a = 0$ we obtain the raw moments

$$\hat{\mu}_r \;=\; \alpha_r \;=\; \mathrm{E}(\mathcal{X}^r) \tag{2.5}$$

whereas $a = \mathrm{E}(\mathcal{X})$ yields the centered moments. The general expressions for the raw r-th moments and centered moments as derived from the density $f(u)$ are

$$\mathrm{E}(\mathcal{X}^r) \;=\; \hat{\mu}_r(\mathcal{X}) \;=\; \int_{-\infty}^{+\infty} u^r \, f(u) \, \mathrm{d}u \quad \text{and} \tag{2.6a}$$

$$\mathrm{E}(\tilde{\mathcal{X}}^r) \;=\; \mu_r(\mathcal{X}) \;=\; \int_{-\infty}^{+\infty} (u - \mu)^r \, f(u) \, \mathrm{d}u \;. \tag{2.6b}$$

The second centered moment is called the *variance*, $\mathrm{var}(\mathcal{X})$ or $\sigma^2(\mathcal{X})$, and its positive square root is the *standard deviation* $\sigma(\mathcal{X})$. The variance is always a non-negative quantity as can be easily shown:

---

[2] Since the moments centered around the expectation value will be used more frequently than the raw moments, we denote them by μ and the raw moments by $\hat{\mu}$. The r-th moment of a distribution is also called the moment of order r.

$$\begin{aligned} \mathrm{var}(\mathcal{X}) \;&=\; \mathrm{E}(\tilde{\mathcal{X}}^2) \;=\; \mathrm{E}\left(\Big(\mathcal{X} - \mathrm{E}(\mathcal{X})\Big)^2\right) \;= \\ &=\; \mathrm{E}\Big(\mathcal{X}^2 \,-\, 2\,\mathcal{X}\,\mathrm{E}(\mathcal{X}) \,+\, \mathrm{E}(\mathcal{X})^2\Big) \;= \\ &=\; \mathrm{E}(\mathcal{X}^2) \,-\, 2\,\mathrm{E}(\mathcal{X})\,\mathrm{E}(\mathcal{X}) \,+\, \mathrm{E}(\mathcal{X})^2 \;= \\ &=\; \mathrm{E}(\mathcal{X}^2) \,-\, \mathrm{E}(\mathcal{X})^2 \;. \end{aligned} \tag{2.7}$$

If $\mathrm{E}(\mathcal{X}^2)$ is finite, than $\mathrm{E}(|\mathcal{X}|)$ is finite too and fulfils the inequality

$$\mathrm{E}(|\mathcal{X}|)^2 \;\leq\; \mathrm{E}(\mathcal{X}^2) \;,$$

and since $\mathrm{E}(\mathcal{X}) \leq \mathrm{E}(|\mathcal{X}|)$ the variance is necessarily a non-negative quantity, $\mathrm{var}(\mathcal{X}) \geq 0$.

If $\mathcal{X}$ and $\mathcal{Y}$ are independent and have finite variances, then we obtain

$$\mathrm{var}(\mathcal{X} \,+\, \mathcal{Y}) \;=\; \mathrm{var}(\mathcal{X}) \,+\, \mathrm{var}(\mathcal{Y}) \;,$$

as follows readily by simple calculation:

$$\begin{aligned} \mathrm{E}\big((\tilde{\mathcal{X}} + \tilde{\mathcal{Y}})^2\big) \;&=\; \mathrm{E}\big(\tilde{\mathcal{X}}^2 \,+\, 2\,\tilde{\mathcal{X}}\,\tilde{\mathcal{Y}} \,+\, \tilde{\mathcal{Y}}^2\big) \;= \\ &=\; \mathrm{E}\big(\tilde{\mathcal{X}}^2\big) \,+\, 2\,\mathrm{E}(\tilde{\mathcal{X}})\,\mathrm{E}(\tilde{\mathcal{Y}}) \,+\, \mathrm{E}\big(\tilde{\mathcal{Y}}^2\big) \;=\; \mathrm{E}\big(\tilde{\mathcal{X}}^2\big) \,+\, \mathrm{E}\big(\tilde{\mathcal{Y}}^2\big) \;, \end{aligned}$$

where we use the fact of vanishing first centered moments: $\mathrm{E}(\tilde{\mathcal{X}}) = \mathrm{E}(\tilde{\mathcal{Y}}) = 0$.

For two general – non necessarily independent – random variables $\mathcal{X}$ and $\mathcal{Y}$, the Cauchy-Schwarz inequality holds for the mixed expectation value:

$$\mathrm{E}(\mathcal{X}\mathcal{Y})^2 \;\leq\; \mathrm{E}(\mathcal{X}^2)\,\mathrm{E}(\mathcal{Y}^2) \;. \tag{2.8}$$

If both random variables have finite variances, the *covariance* is defined by

$$\begin{aligned} \mathrm{cov}(\mathcal{X},\mathcal{Y}) \;&=\; \mathrm{var}(\mathcal{X},\mathcal{Y}) \;=\; \mathrm{E}\Big(\big(\mathcal{X} - \mathrm{E}(\mathcal{X})\big)\big(\mathcal{Y} - \mathrm{E}(\mathcal{Y})\big)\Big) \;= \\ &=\; \mathrm{E}\Big(\mathcal{X}\mathcal{Y} \,-\, \mathcal{X}\,\mathrm{E}(\mathcal{Y}) \,-\, \mathrm{E}(\mathcal{X})\,\mathcal{Y} \,+\, \mathrm{E}(\mathcal{X})\,\mathrm{E}(\mathcal{Y})\Big) \;= \\ &=\; \mathrm{E}(\mathcal{X}\mathcal{Y}) \,-\, \mathrm{E}(\mathcal{X})\,\mathrm{E}(\mathcal{Y}) \;. \end{aligned} \tag{2.9}$$

The covariance $\mathrm{cov}(\mathcal{X},\mathcal{Y})$ and the *coefficient of correlation* $\rho(\mathcal{X},\mathcal{Y})$,

$$\mathrm{cov}(\mathcal{X},\mathcal{Y}) \;=\; \mathrm{E}(\mathcal{X}\mathcal{Y}) \,-\, \mathrm{E}(\mathcal{X})\,\mathrm{E}(\mathcal{Y}) \quad\text{and}\quad \rho(\mathcal{X},\mathcal{Y}) \;=\; \frac{\mathrm{cov}(\mathcal{X},\mathcal{Y})}{\sigma(\mathcal{X})\,\sigma(\mathcal{Y})} \;, \tag{2.9'}$$

are a measure of correlation between the two variables. As a consequence of the Cauchy-Schwarz inequality we have $-1 \leq \rho(\mathcal{X},\mathcal{Y}) \leq 1$. If covariance and correlation coefficient are equal to zero, the two random variables $\mathcal{X}$ and $\mathcal{Y}$

are *uncorrelated*. Independence implies lack of correlation but in general is the stronger property (section 2.3.4).

In addition to the expectation value two more quantities are used to characterize the center of probability distributions (figure 2.1): (i) The *median* $\bar{\mu}$ is the value at which the number of points of a distribution at lower values of matches exactly the number of points at higher values as expressed in terms of two inequalities,

$$\begin{aligned} P(\mathcal{X} \leq \bar{\mu}) \;\geq\; \frac{1}{2} \;\text{ and }\; P(\mathcal{X} \geq \bar{\mu}) \;\geq\; \frac{1}{2} \;\text{ or} \\ \int_{-\infty}^{\bar{\mu}} dF(x) \;\geq\; \frac{1}{2} \;\text{ and }\; \int_{\bar{\mu}}^{+\infty} dF(x) \;\geq\; \frac{1}{2}\;, \end{aligned} \tag{2.10}$$

where Lebesgue-Stieltjes integration is applied or in case of an absolutely continuous distribution the condition simplifies to

$$P(\mathcal{X} \leq \bar{\mu}) \;=\; P(\mathcal{X} \geq \bar{\mu}) \;=\; \int_{-\infty}^{\bar{\mu}} f(x)\,\mathrm{d}x \;=\; \frac{1}{2}\;, \tag{2.10'}$$

and (ii) the *mode* $\widetilde{\mu}$ of a distribution is the most frequent value – the value that is most likely obtained through sampling – and it coincides with the maximum of the probability mass function for discrete distribution or the maximum of the probability density in the continuous case. An illustrative example for the discrete case is the probability mass function of the scores for throwing two dice, where the mode is $\widetilde{\mu} = 7$ (figure 1.11). A probability distribution may have more than one mode. Bimodal distributions occur occasionally and then the two modes provide much more information on the expected outcomes than mean or median (subsection 2.5.8).

Median and mean are related by an inequality, which says that the difference between both is bounded by one standard deviation [302, 326]:

$$\begin{aligned} |\mu - \overline{\mu}| \;&=\; |\mathrm{E}(\mathcal{X} - \overline{\mu}| \;\leq\; \mathrm{E}(|\mathcal{X} - \overline{\mu}|) \;\leq \\ &\leq\; \mathrm{E}(|\mathcal{X} - \mu|) \;\leq\; \sqrt{\mathrm{E}\Big((\mathcal{X} - \mu)^2\Big)} \;=\; \sigma\;. \end{aligned} \tag{2.11}$$

The absolute difference between mean and median can't be larger than one standard deviation of the distribution.

For many purposes a generalization of the median from two to n equally sized data sets is useful. The *quantiles* are points taken at regular intervals from the cumulative distribution function $F(x)$ of a random variable $\mathcal{X}$. Ordered data are divided into n essentially equal-sized subsets and accordingly, $(n-1)$ points on the x-axis separate the subsets. Then, the k-th n-quantile is defined by $P(\mathcal{X} < x) \leq \frac{k}{n} = p$ (figure 2.2) or equivalently

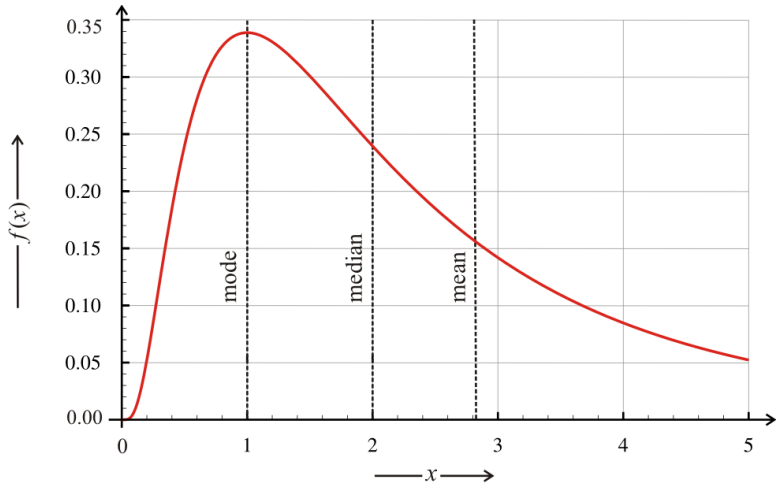


Fig. 2.1 Probability densities and moments. As an example of an asymmetric distribution with highly different values for mode, median, and mean, the lognormal density

$$f(x) = \tfrac{1}{\sqrt{2\pi}\,\sigma\,x}\exp\bigl(-(\ln x-\nu)^2/(2\sigma^2)\bigr)$$

is shown. Parameters values: $\sigma = \sqrt{\ln 2}$, $\nu = \ln 2$ yielding $\widetilde{\mu} = \exp(\nu - \sigma^2/2) = 1$ for the mode, $\bar{\mu} = \exp(\nu) = 2$ for the median and $\mu = \exp(\nu+\sigma^2/2) = 2\sqrt{2}$ for the mean, respectively. The sequence mode < median < mean is characteristic for distributions with positive skewness whereas the opposite sequence mean < median < mode is found in cases of negative skewness (see also figure 2.3).

$$F^{-1}(p) \doteq \inf\{x \in \mathbb{R} : F(x) \geq p\} \quad \text{and} \quad p = \int_{-\infty}^{x} \mathrm{d}F(u)\ . \tag{2.12}$$

In case the random variable has a probability density the integral simplifies to $p = \int_{-\infty}^{x} f(u)\,\mathrm{d}u$. The median is simply the value of x for $p = \frac{1}{2}$. For partitioning into four parts we haver the first or lower quartile at $p = \frac{1}{4}$, the second quartile or median at $p = \frac{1}{2}$, and the third or upper quartile at $p = \frac{3}{4}$. The lower quartile contains 25 % of the data, the median 50 %, and the upper quartile eventually 75 % of the data.

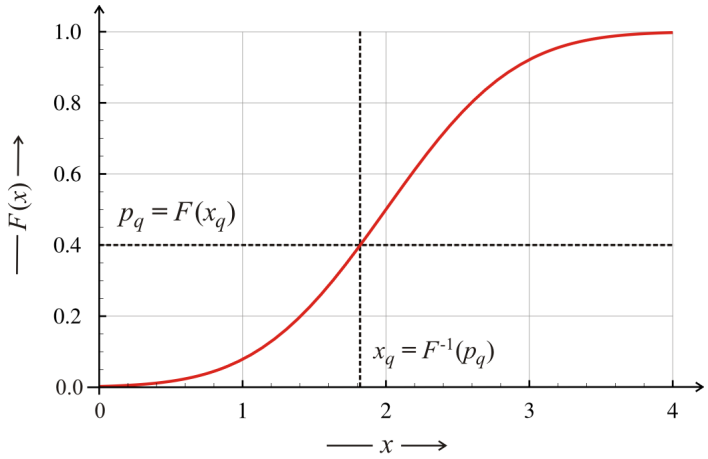


Fig. 2.2 Definition and determination of quantiles. A quantile q with $p_q = k/n$ defines a value x_q at which the (cumulative) probability distribution reaches the value $F(x_q) = p_q$ corresponding to $P(\mathcal{X} < x) \leq p$. The figure shows how the position of the quantile $p_q = k/n$ is used to determine its value $x_q(p)$. In particular we use here the normal distribution $\mathcal{N}(x)$ as function $F(x)$ and the computation yields $F(x_q) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x_q - \nu}{\sqrt{2\sigma^2}}\right)\right) = p_q$. Parameter choice: $\nu = 2$, $\sigma^2 = \frac{1}{2}$, and for the quantile $(n = 5, k = 2)$, yielding $p_q = 2/5$ and $x_q = 1.8209$.

2.1.2 Higher moments

Two other quantities related to higher moments are frequently used for a more detailed characterization of probability distributions:[3] (i) The *skewness*

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3}{\sigma^3} = \frac{\mathrm{E}\Big(\big(\mathcal{X} - \mathrm{E}(\mathcal{X})\big)^3\Big)}{\Big(\mathrm{E}\Big(\big(\mathcal{X} - \mathrm{E}(\mathcal{X})\big)^2\Big)\Big)^{3/2}} \tag{2.13}$$

and (ii) *kurtosis*

$$\begin{aligned} \beta_2 &= \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4} = \frac{\mathrm{E}\Big(\big(\mathcal{X} - \mathrm{E}(\mathcal{X})\big)^4\Big)}{\Big(\mathrm{E}\Big(\big(\mathcal{X} - \mathrm{E}(\mathcal{X})\big)^2\Big)\Big)^2} \quad \text{and} \\ \gamma_2 &= \frac{\kappa_4}{\kappa_2^2} = \frac{\mu_4}{\sigma^4} - 3 = \beta_2 - 3\,, \end{aligned} \tag{2.14}$$

[3] In contrast to expectation value, variance and standard deviation, skewness and kurtosis are not uniquely defined and it is necessary therefore to check carefully the author's definitions when reading text from literature.

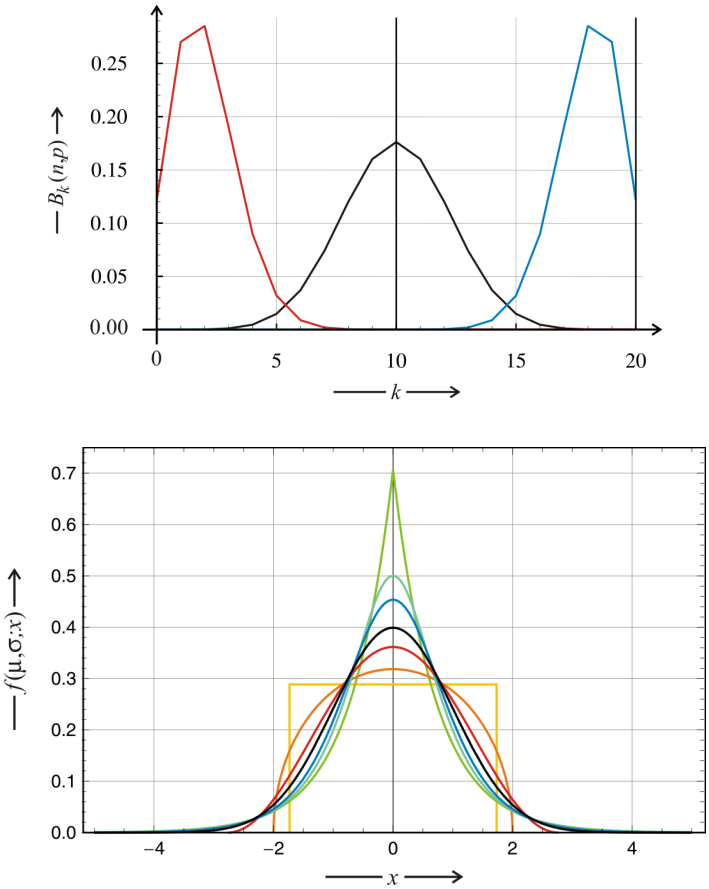


Fig. 2.3 Skewness and kurtosis. The upper part of the figures illustrates the sign of skewness with asymmetric density functions. The examples are taken form the binomial distribution $B_k(n,p)$: $\gamma_1 = (1-2p)/\sqrt{np\,(1-p)}$ with $p = 0.1$ (red), 0.5 (black; symmetric), and 0.9 (blue) with the values $\gamma_1 = 0.596$, 0, -0.596.
Densities with different kurtosis are compared in the lower part of the figure: The Laplace distribution (chartreuse), the hyperbolic secant distribution (green), and the logistic distribution (blue) are leptokurtic with excess kurtosis values 3, 2, and 1.2, respectively. The normal distribution is the reference curve with excess kurtosis 0 (black). The raised cosine distribution (red), the Wigner semicircle distribution (orange), and the uniform distribution (yellow) are platykurtic with excess kurtosis values of -0.593762, -1, and -1.2 respectively. All densities are calibrated such that $\mu = 0$ and $\sigma^2 = 1$ (The picture is recalculated and redrawn from `http://en.wikipedia.org/wiki/Kurtosis`, March 30,2011).

which is either defined as the fourth standardized moment β_2 or in terms of cumulants κ_n given as *excess kurtosis*, γ_2.

Skewness is a measure of the asymmetry of the probability density: curves that are symmetric about the mean have zero skew, *negative skew* implies a longer left tail of the distribution caused by more low values, and *positive skew* is characteristic for a distribution with a longer right tail. Positive skew is quite common with empirical data (see, for example the log-normal distribution in section 2.5.1).

Kurtosis characterizes the degree of *peakedness* of a distribution. High kurtosis implies a sharper peak and flat tails, low kurtosis in contrary characterizes flat or round peaks and thin tails. Distributions are called *leptokurtic* if they have a positive excess kurtosis and therefore are sharper peak and a thicker tail than the *normal distribution* (section 2.3.3), which is taken as a reference with zero kurtosis. Distributions are characterized as *platykurtic* if they have a negative excess kurtosis, a broader peak and thinner tails. In figure 2.3 the following seven distributions, standardized to $\mu = 0$ and $\sigma^2 = 1$, are compared there with respect to kurtosis:

(i) Laplace distribution: $f(x) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$, $b = \frac{1}{\sqrt{2}}$,
(ii) hyperbolic secant distribution: $f(x) = \frac{1}{2} \operatorname{sech}\left(\frac{\pi}{2} x\right)$,
(iii) logistic distribution: $f(x) = \frac{e^{-(x-\mu)/s}}{s\,(1+e^{-(x-\mu)/s})^2}$, $s = \sqrt{3}/\pi$,
(iv) normal distribution: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/(2\sigma^2)}$,
(v) raised cosine distribution: $f(x) = \frac{1}{2s}\Big(1+\cos(\pi(x-\mu)/s)\Big)$, $s = \frac{1}{\sqrt{\frac{1}{3}-\frac{2}{\pi^2}}}$,
(vi) Wigner's semicircle: $f(x) = \frac{2}{\pi r^2}\sqrt{r^2 - x^2}$, $r = 2$, and
(vii) uniform distribution: $f(x) = \frac{1}{b-a}$, $b - a = 2\sqrt{3}$.

These seven functions span the whole range of maxima from a sharp peak to a completely flat plateau with the normal distribution chosen as reference function (figure 2.3). Distribution (i), (ii), and (iii) are leptokurtic whereas (v), (vi), and (vii) are platykurtic.

One property of skewness and kurtosis being caused by definition is important to note: The expectation value, the standard deviation, and the variance are quantities with dimensions, whereas skewness and kurtosis are defined as dimensionless numbers.

The cumulants κ_n are the coefficients of a series expansion of the logarithm of the *characteristic function* (2.29), which in turn is the Fourier transform of the probability density function, $f(x)$, or the logarithm of the *moment generating function* (2.27) as discussed in section 2.2:

$$\begin{aligned} h(s) &= \ln \phi(s) = \sum_{n=1}^{\infty} \kappa_n \frac{(\dot{u}\, s)^n}{n!} \text{ with} \\ \phi(s) &= \int_{-\infty}^{+\infty} \exp(\dot{u}\, s\, x)\, f(x)\, \mathrm{d}x \ . \end{aligned} \tag{2.15}$$

The first five cumulants κ_n $(n = 1, \ldots, 5)$ expressed in terms of the expectation value μ and the central moments μ_n $(\mu_1 = 0)$ are

$$\begin{aligned} \kappa_1 &= \mu \\ \kappa_2 &= \mu_2 \\ \kappa_3 &= \mu_3 \\ \kappa_4 &= \mu_4 - 3\,\mu_2^2 \\ \kappa_5 &= \mu_5 - 10\,\mu_2\mu_3 \ . \end{aligned} \tag{2.16}$$

We shall come back to the use of cumulants κ_n in sections 2.3 and 2.5 on the occasion of a comparison of frequently used individual probability densities and in section 2.6 when we apply k-statistics in order to compute empirical moments from incomplete data sets.

2.1.3 Information entropy

Information theory has been developed during World War Two as theory of the communication of secret messages. No wonder that the theory was conceived and worked out at Bell Labs and the person who was the leading figures in this area was an American cryptographer, electronic engineer and computer scientist, Claude Elwood Shannon [405, 406]. One of the central

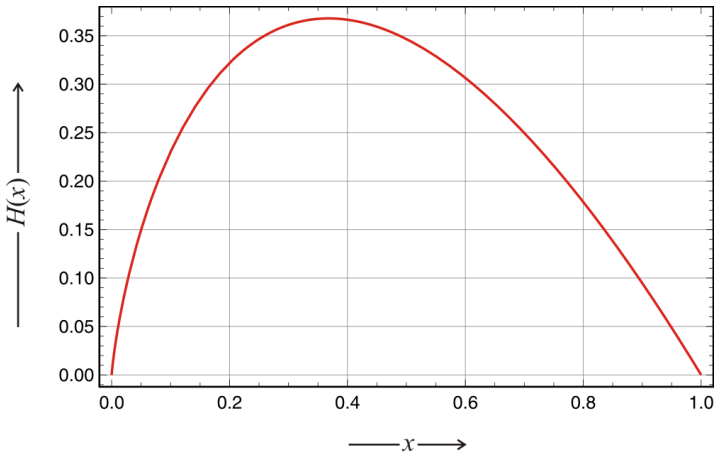


Fig. 2.4 The functional relation of information entropy. The plot shows the function $H = -\,x\,\ln x$ in the range $0 \leq x \leq 1$. For $x = 0$ the convention of probability theory, $-\,0\ln 0 = 0 \cdot \infty = 0$, is applied.

issues of information theory is *self-information* or the *content of information*

$$I(\omega) \;=\; \mathrm{lb}\left(\frac{1}{P(\omega)}\right) \;=\; -\,\mathrm{lb}\,P(\omega) \tag{2.17}$$

that can be encoded, for example, in a sequence of given length. Commonly one thinks about binary sequences and therefore the information is measured in *binary digits* or *bits*.[4] The rationale behind this expression is the definition of a measure of information that is positive and additive for independent events. From equation (1.34) follows:

$$P(AB) \;=\; P(A)\cdot P(B) \;\;\Longrightarrow\;\; I(A\cap B) \;=\; I(AB) \;=\; I(A) + I(B)\;,$$

and this relation is fulfilled by the logarithm. Since $P(\omega) \leq 1$ by definition, the negative logarithm is a positive quantity. Equation (2.17) yields zero information for an event taking place with certainty, $P(\omega) = 1$. The outcome of the fair coin toss with $P(\omega) = \frac{1}{2}$ provides 1 bit information, and rolling two 'six' with two dice in one throw has a probability $P(\omega) = \frac{1}{36}$ and yields 5.17 bits (For a modern treatise of information theory and entropy see [179]).

Countable sample space. In order to measure the information content of a probability distribution Claude Shannon introduced the information entropy, which is simply the expectation value of the information content and which is represented by a function that resembles the expression for the thermodynamic entropy in statistical mechanics. We consider first the discrete case of a probability mass function $p_k = P(\mathcal{X} = x_k)\,,\; k \in \mathbb{N}_{>0}, k \leq n$:

$$H(p) \;=\; -\sum_{k=1}^{n} p_k \,\log\, p_k \;\;\text{ with }\;\; p_k \geq 0\,, \sum_{k=1}^{n} p_k = 1\;. \tag{2.18}$$

Thus, the entropy can be visualized as the expectation value of the negative logarithm of the probabilities

$$H(p) \;=\; \mathrm{E}(-\log\, p_k) \;=\; \mathrm{E}\left(\log\Big(\frac{1}{p_k}\Big)\right)\;,$$

where the term $\log(1/p_k)$ can be viewed as the number of bits to be assigned to to the point x_k provided the binary logarithm is used ($\log \equiv \mathrm{lb}$).

The functional relationship, $H = -x\log x$, on the interval $0 \leq x \leq 1$ underlying the information entropy is a concave function (figure 2.4). It is easily shown that the entropy of a discrete probability distribution is always

[4] The logarithm is taken to the base 2 and it is commonly called *binary logarithm* or *logarithmus dualis*: $\log_2 \equiv \mathrm{lb} \equiv \mathrm{ld}$. In informatics the conventional unit of information is the *byte*: 1 byte (B) = 8 bits being tantamount to the coding capacity of an eight digit binary sequence. Although there is little chance of confusion, one should be aware that in the International System of Units 'B' is the abbreviation for the acoustical unit 'bel', which is the unit for measuring the signal strength of sound.

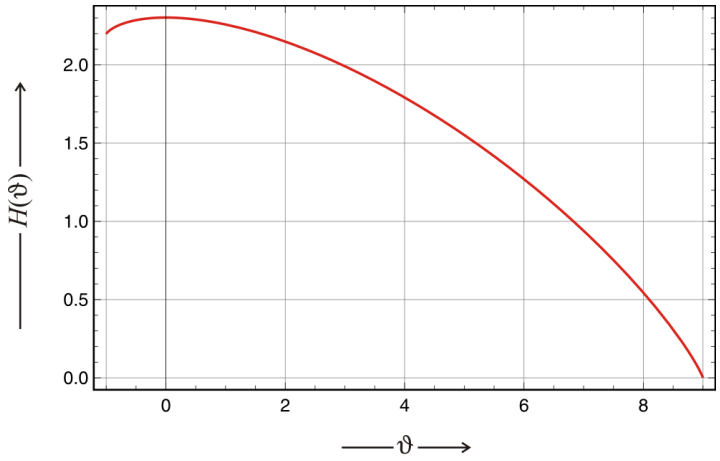


Fig. 2.5 Maximum information entropy. The discrete probability distribution with maximal information entropy in the uniform distribution $\mathcal{U}_p$. The entropy of the probability distribution $p_1 = \frac{1-\vartheta}{n}$ and $p_j = (1 - \frac{\vartheta}{n-1})/n \, \forall \, j = 2, 3, \ldots, n$ with $n = 10$ is plotted against the parameter ϑ. All probabilities p_k are defined and the entropy $H(\vartheta)$ is real and non-negative on the interval $-1 \leq \vartheta \leq 9$ and passes a maximum at $\vartheta = 0$.

non-negative. A verification of this conjecture can be given, for example, by considering the two extreme cases: (i) there almost certainly only one outcome, $p_1 = P(\mathcal{X} = x_1) = 1$ and $p_j = P(\mathcal{X} = x_j) = 0 \; \forall \; j \in \mathbb{N}_{>0}, j \neq 1$, and the information entropy $H = 0$ in this completely determined case, and (ii) all events have the same probability, we are dealing with the uniform distribution, $p_k = P(\mathcal{X} = x_k) = \frac{1}{n}$, or a case of the principle of indifference, the entropy is positive, and takes on its maximum value, $H(p) = \log n$. The entropies of all other discrete distributions lie in between:

$$0 \leq H(p) \leq \log n \quad \text{or} \quad H(p) \leq \log \, n \, , \tag{2.19}$$

and the value of the entropy is a measure of the lack of information on the distribution. Case (i) is deterministic and we have the full information on the outcome *a priori*, whereas case (ii) provides maximal uncertainty because all outcomes have the same probability. A rigorous proof that the uniform distribution has maximum information entropy among all discrete distributions is found in the literature [71, 75]. We dispense from reproducing the proof here but we illustrate by means of figure 2.5: The starting point is the uniform distribution of n events with a probability of $p = \frac{1}{n}$ for each one, and then we attribute a different probability to a single event: $p_1 = \frac{1-\vartheta}{n}$

Table 2.1 Probability distributions with maximum information entropy. The table compares three probability distributions with maximum entropy: (i) the discrete uniform distribution on the support $\Omega = \{1 \leq k \leq n, k \in \mathbb{N}\}$, (ii) the exponential distribution on $\Omega = \mathbb{R}_{\geq 0}$, and the normal distribution on $\Omega = \mathbb{R}$.

Distribution	Space Ω	Density	Mean	Var	Entropy
uniform	$\mathbb{N}_{>0}$	$\frac{1}{n} \; \forall \; k = 1, \ldots, n$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\log \, n$
exponential	$\mathbb{R}_{\geq 0}$	$\frac{1}{\mu} \, e^{-x/\mu}$	μ	μ^2	$1 + \log \, \mu$
normal	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\frac{(x-\mu^2)}{2\sigma^2}}$	μ	σ^2	$\Big(1 + \log(2\pi\sigma^2)\Big)/2$

and $p_j = (1 - \frac{\vartheta}{n-1})/n$ $(j = 2, 3, \ldots, n)$. The entropy of the distribution is considered as a function of ϑ and indeed the maximum occurs at $\vartheta = 0$.

Uncountable measurable sample space. The information entropy of a continuous probability density $p\,(x)$ with $x \in \mathbb{R}$ is calculated by means of integration

$$H(p) \;=\; -\int_{-\infty}^{+\infty} p\,(x) \, \log \, p\,(x) \, \mathrm{d}x \quad \text{with} \quad p_k \geq 0\,, \int_{-\infty}^{+\infty} p\,(x) \, \mathrm{d}x = 1\,, \tag{2.18'}$$

and as in the discrete case we can write the entropy as an expectation value of $\log(1/p)$:

$$H(p) \;=\; \mathrm{E}\Big(-\log \, p\,(x)\Big) \;=\; \mathrm{E}\left(\log\Big(\frac{1}{p\,(x)}\Big)\right) \;.$$

We consider two specific examples that are distributions with maximum entropy: the exponential distribution (section 2.5.4) on $\Omega = \mathbb{R}_{\geq 0}$ with the density

$$f_{\exp}(x) \;=\; \frac{1}{\mu} \, e^{-\frac{x}{\mu}} \;,$$

the mean μ, and the variance μ^2, and the normal distribution (section 2.3.3) on $\Omega = \mathbb{R}$ with the density

$$f_{\mathcal{N}}(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \;,$$

the mean μ, and the variance σ^2

In the discrete case we made a seemingly unconstrained search for the distribution of maximum entropy, although the discrete uniform distribution contained the number of sample points n as input restriction and indeed, n appears as parameter in the analytical expression for the entropy (table 2.1). Now, in the continuous case the constraints become more evident since we

shall use fixed mean (μ) or fixed mean (μ) and variance (σ^2) as the basis of comparison in the search for distributions with maximum entropy.

The entropy of the exponential density on the sample space $\Omega = \mathbb{R}_{\geq 0}$ with mean μ and variance μ^2 is calculated to be

$$H(f_{\exp}) \;=\; -\int_0^\infty \frac{1}{\mu}\, e^{-x/\mu} \left(-\log\, \mu \,-\, \frac{x}{\mu} \right) \mathrm{d}x \;=\; 1 \,+\, \log\, \mu\,. \qquad (2.20)$$

In contrast to the discrete case the entropy of the exponential probability density can become negative for small μ-values as can be easily visualized by considering the shape of the density: Since $\lim_{x\to 0} f_{\exp}(x) = 1/\mu$, an appreciable fraction of the density function adopts values $f_{\exp}(x) > 1$ for sufficiently small μ and then $-p \log p < 0$ is negative. Among all continuous probability distributions with mean $\mu > 0$ on the support $\mathbb{R}_{\geq 0} = [0, \infty[$ the exponential distribution has the maximum entropy. Proofs for this conjecture are available in the literature [71, 75, 357].

For the normal density we obtain from equation (2.18'):

$$\begin{aligned} H(f_{\mathcal{N}}) \;&=\; -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left(-\log(\sqrt{2\pi\sigma^2}) - \frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2 \right) \mathrm{d}x \\ &=\; \frac{1}{2}\Big(1 + \log(2\pi\sigma^2)\Big)\,. \end{aligned} \qquad (2.21)$$

It is not unexpected that the information entropy of the normal distribution is independent of the mean μ, which causes nothing but a shift of the whole distribution along the x-axis: all Gaussian densities with the same variance σ^2 have the same entropy. Again we see that the entropy of the normal probability density can become negative for sufficiently small values of σ^2. The normal distribution is distinguished among all continuous distributions on $\Omega = \mathbb{R}$ with given variance σ^2, since it is the distribution with maximum entropy. Several proofs for this theorem were developed, we refer again to the literature [71, 75, 357]. The three distributions with maximum entropy are compared in table 2.1.

Principle of maximum entropy. The information entropy can be interpreted as the required amount of information we would need in order to fully describe the system. Equations (2.18) and (2.18') are the basis of a search for probability distribution with maximum entropy under certain constraints, for example constant mean μ or constant variance σ^2. The maximum entropy principle has been introduced by the American physicist Edwin Thompson Jaynes as a method of statistical inference [232, 233]: He suggests to use those probability distributions, which satisfy the prescribed constraints and have the largest entropy. The rationale for this choice is to use a probability distribution that reflects our knowledge and does not contain any unwarranted information. The predictions made on the basis of a probability distribution with maximum entropy should be least surprising. If we chose a distribution with smaller entropy, this distribution would contain more information than

justified by our *a priori* understanding of the problem. It is useful to illustrate a typical strategy [71]: *"..., the principle of maximum entropy guides us to the best probability distribution that reflects our current knowledge and it tells us what to do if experimental data do not agree with predictions coming from our chosen distribution: Understand why the phenomenon being studied behaves in an unexpected way, find a previously unseen constraint, and maximize the entropy over the distributions that satisfy all constraints we are now aware of, including the new one."* We realize a different way of thinking about probability that becomes even more evident in Bayesian statistics, which is sketched in sections 1.3 and 2.6.4.

The choice of the word *entropy* for the expected information content of a distribution is not accidental. Ludwig Boltzmann's statistical formula[5]

$$S \;=\; k_{\mathrm{B}} \,\ln W \quad \text{with} \quad W \;=\; \frac{N!}{N_1!N_2!\cdots N_m!}\;, \tag{2.22}$$

with W being the so-called thermodynamic probability and k_{B} Boltzmann's constant: $k_{\mathrm{B}} = 1.38065 \times 10^{-23}\,\mathrm{Joule{\cdot}Kelvin}^{-1}$ and $N = \sum_{j=1}^{m} N_j$ being the total number of particles being distributed over m states with the frequencies $p_k = N_k/N$ and $\sum_{j=1}^{m} p_j = 1$. The number of particles N is commonly very large and Stirling's formula named after the Scottish mathematician James Stirling applies: $n! \approx n \ln n$, and this leads to:

$$S \;=\; k_{\mathrm{B}} \left(N \,\ln N - \sum_{i=1}^{m} N_i \,\ln N_i \right) \;=\; -\,k_{\mathrm{B}}\, N \left(-\,\ln N + \sum_{i=1}^{m} \frac{N_i}{N} \,\ln N_i \right) \;=$$

$$=\; -\,k_{\mathrm{B}}\, N \sum_{i=1}^{m} p_i \,\ln p_i \;.$$

For a single particle we obtain an entropy

$$s \;=\; \frac{S}{N} \;=\; -\,k_{\mathrm{B}} \sum_{i=1}^{m} p_i \,\ln p_i \;, \tag{2.22'}$$

which is identical with Shannon's formula (2.18) except the factor containing the universal constant k_{B}.

Eventually we point at some important differences between thermodynamic entropy and information entropy that should be kept in mind when discussing analogies between them. The thermodynamic principle of maximum entropy is a physical law known as the second law of thermodynamics:

[5] A few remarks are important: Equation (2.22) is Max Planck's expression for the entropy in statistical mechanics, although it has been carved in Boltzmann's tomb stone, and W is called a probability despite the fact that it is not normalized, $W \geq 0$.

The entropy of an isolated system[6] is nondecreasing in general and increasing if processes are taking place, and hence approaches a maximum. The principle of maximum entropy in statistics is a rule for appropriate design of distribution functions and has the rank of a guideline and not that of a natural law. Thermodynamic entropy is an *extensive property* and this means that it increases with the size of the system. Information entropy, on the other hand, is an *intensive property* and insensitive to size. An illustrative example of this difference is due to the Russian biophysicist Mikhail Vladimirovich Volkenshtein [451]: Considering the process of flipping a coin in reality and calculating all contributions to the process shows that the information entropy is a minute contribution to the thermodynamic entropy only. The total thermodynamic entropy change as a result of the coin flipping process is dominated by far by the metabolic contributions of the flipping individual, as there are muscle contraction, joint rotations, and by the heat production on the surface where the coin lands, etc. Imagine the thermodynamic entropy production if you flip a coin two meters high – the gain in information remains still one bit!

[6] A isolated system exchanges neither matter nor energy with its environment (For isolated, closed, and open systems see also section 4.3).

2.2 Generating functions

In this section we introduce auxiliary functions, which allow for the derivation of compact representations of probability distributions and which provide convenient tools for handling functions of probabilities. The generating functions commonly contain one or more auxiliary variables – here denoted by s, which are lacking direct physical meaning but enable straightforward calculation of properties of random variables at certain values of s. In particular we shall make use of probability generating functions $g(s)$, moment generating functions $M(s)$ and characteristic functions $\phi(s)$. The characteristic function $\phi(s)$ exists for all distributions but we shall encounter cases where no probability or moment generating functions exist (see, for example, the Cauchy-Lorentz distribution in subsection 2.5.6). In addition to the three generating functions mentioned here other functions are in use as well. An example is the *cumulant generating function* that is lacking a uniform definition. It is either the logarithm of the moment generating function or the logarithm of the characteristic function – we shall mention both.

2.2.1 Probability generating functions

Let $\mathcal{X}$ be a random variable taking only non-negative integer values with a probability distribution given by

$$P(\mathcal{X} = j) \;=\; a_j \,;\; j = 0, 1, 2, \ldots \;. \tag{2.23}$$

A auxiliary variable s is introduced and the *probability generating function* is expressed by an infinite power series

$$g(s) \;=\; a_0 \,+\, a_1\, s \,+\, a_2\, s^2 \,+\, \ldots \;=\; \sum_{j=0}^{\infty} a_j\, s^j \;=\; \mathrm{E}(s^{\mathcal{X}}) \;. \tag{2.24}$$

As we shall show later, the full information on the probability distribution is encapsulated in the coefficients a_j ($j \in \mathbb{N}$). Intuitively this is no surprise since the coefficients a_j are the individual probabilities of a probability mass function in (1.28'): $a_j = p_j$. The expression of the probability generation function as an expectation value is useful in the comparison with other generating functions.

In most cases s is a real valued variable, although it can be of advantage to consider also complex s. Recalling $\sum_j a_j = 1$ from (2.23) we verify easily that the power series (2.24) converges for $|s| \leq 1$:

$$|g(s)| \;\leq\; \sum_{j=0}^{\infty} |a_j| \cdot |s|^j \;\leq\; \sum_{j=0}^{\infty} a_j \;=\; 1 \;, \quad \text{for } |s| \leq 1 \;.$$

For $|s| < 1$ we can differentiate[7] the series term by term in order to calculate the derivatives of the generating function $g(s)$

$$\frac{\mathrm{d}g}{\mathrm{d}s} \;=\; g'(s) \;=\; a_1 \,+\, 2\,a_2 s \,+\, 3\,a_3 s^2 \,+\, \ldots \;=\; \sum_{n=1}^{\infty} n\,a_n s^{n-1} \;,$$

$$\frac{\mathrm{d}^2 g}{\mathrm{d}s^2} \;=\; g''(s) \;=\; 2\,a_2 \,+\, 6\,a_3 s \,+\, \ldots \;=\; \sum_{n=2}^{\infty} n(n-1)\,a_n s^{n-2} \;,$$

and, in general, we have

$$\frac{\mathrm{d}^j g}{\mathrm{d}s^j} \;=\; g^{(j)}(s) \;=\; \sum_{n=j}^{\infty} n(n-1)\ldots(n-j+1)\,a_n s^{n-j} \;=$$

$$= \sum_{n=j}^{\infty} (n)_j \; a_n \, s^{n-j} \;=\; \sum_{n=j}^{\infty} \binom{n}{j} \, j\,!\; a_n \, s^{n-j} \;,$$

where $(x)_n \equiv (x-n+1)^{(n)}$ stands for the falling Pochhammer symbol.[8] Setting $s = 0$, all terms vanish except the constant term

$$\frac{\mathrm{d}^j g}{\mathrm{d}s^j}\,\Big|_{s=0} \;=\; g^{(j)}(0) \;=\; j\,!\,a_j \quad \text{or} \quad a_j \;=\; \frac{1}{j\,!}\,g^{(j)}(0) \;.$$

In this way all a_j's may be obtained by consecutive differentiation from the generating function and alternatively the generating function can be determined from the known probability distribution.

---

[7] Since we need the derivatives very often in this section, we make advantage of short notations: $\mathrm{d}g(s)/\mathrm{d}s = g'(s)$, $\mathrm{d}^2 g(s)/\mathrm{d}s^2 = g''(s)$, and $\mathrm{d}^j g(s)/\mathrm{d}s^j = g^{(j)}(s)$ and for simplicity also $(\mathrm{d}g/\mathrm{d}s)|_{s=k} = g'(k)$ and $(\mathrm{d}^2 g/\mathrm{d}s^2)|_{s=k} = g''(k)$ $(k \in \mathbb{N})$.

[8] The definition of the Pochhammer symbol is ambiguous [256, p. 414]. In combinatorics the Pochhammer symbol $(x)_n$ is used for the *falling factorial*,

$$(x)_n \;=\; x(x-1)(x-2)\cdots(x-n+1) \;=\; \frac{\Gamma(x+1)}{\Gamma(x-n+1)} \;,$$

whereas the *rising factorial* is the written a

$$x^{(n)} \;=\; x(x+1)(x+2)\cdots(x+n-1) \;=\; \frac{\Gamma(x+n)}{\Gamma(x)} \;.$$

We also mention a useful identity between the partial factorials

$$(-x)^{(n)} \;=\; (-1)^n\,(x)_n \;.$$

In the theory of special functions in physics and chemistry, in particular in the context of the hypergeometric functions, however, $(x)_n$ is used for the rising factorial. Here, we shall use both definitions depending on the context but we shall always say whether we mean the rising or the falling factorial. Clearly, the expression in terms of the Gamma function is unambiguous.

Putting $s = 1$ in $g'(s)$ and $g''(s)$ we can compute the first and second moments of the distribution of $\mathcal{X}$:

$$\begin{aligned} g'(1) &= \sum_{n=0}^{\infty} n\, a_n = \mathrm{E}(\mathcal{X}) \,, \\ g''(1) &= \sum_{n=0}^{\infty} n^2\, a_n - \sum_{n=0}^{\infty} n\, a_n = \mathrm{E}(\mathcal{X}^2) - \mathrm{E}(\mathcal{X}) \\ \mathrm{E}(\mathcal{X}) &= g'(1)\,, \quad \text{and} \\ \mathrm{E}(\mathcal{X}^2) &= g'(1) + g''(1) \quad \text{and} \quad \mathrm{var}(\mathcal{X}) = g'(1) + g''(1) - g'(1)^2 \,. \end{aligned} \tag{2.25}$$

We summarize: The probability distribution of a non-negative integer values random variable can be converted into a generating function without loosing information. The generating function is uniquely determined by the distribution and *vice versa*.

2.2.2 Moment generating functions

Basis of the moment generating function is the series expansion of the exponential function of the random variable $\mathcal{X}$

$$e^{\mathcal{X}s} = 1 + \mathcal{X}\, s + \frac{\mathcal{X}^2}{2!}\, s^2 + \frac{\mathcal{X}^3}{3!}\, s^3 \ldots \,.$$

The *moment generating function* allows for direct computation of the moments of a probability distribution as defined in equation (2.23) since we have:

$$M_{\mathcal{X}}(s) = \mathrm{E}(e^{\mathcal{X}\, s}) = 1 + \hat{\mu}_1\, s + \frac{\hat{\mu}_2}{2!}\, s^2 + \frac{\hat{\mu}_3}{3!}\, s^3 \ldots = 1 + \sum_{n=1}^{\infty} \hat{\mu}_n \frac{s^n}{n!} \,. \tag{2.26}$$

wherein $\hat{\mu}_i$ is the i-th raw moment. The moments are obtained by differentiating $M_{\mathcal{X}}(s)$ with respect to s and then setting $s = 0$. From the n-th derivative we obtain

$$\mathrm{E}(\mathcal{X}^n) = \hat{\mu}_n = M_{\mathcal{X}}^{(n)} = \frac{\mathrm{d}^n M_{\mathcal{X}}}{\mathrm{d}s^n}\Big|_{s=0} \,.$$

A probability distribution thus has (at least) as many moments as many times the moment generating function can be continuously differentiated (see also characteristic function in subsection 2.2.3). If two distributions have the same moment generating functions they are identical at all points:

$$M_{\mathcal{X}}(s) = M_{\mathcal{Y}}(s) \iff F_{\mathcal{X}}(x) = F_{\mathcal{Y}}(x) \,.$$

This statement, however, does not imply that two distributions are identical when they have the same moments, because in some cases the moments exist but the moment generating function does not, since the limit $\lim_{n\to\infty} \sum_{k=0}^{n} \frac{\hat{\mu}_k\, s^k}{k!}$ diverges as, for example, in case of the logarithmic normal distribution.

The real *cumulant generating function* is the formal logarithm of the moment generating function that can be expanded in a power series

$$\begin{aligned} k(s) \;&=\; \ln\Big(\mathrm{E}\big(e^{\mathcal{X}s}\big)\Big) \;=\; -\sum_{n=1}^{\infty} \frac{1}{n}\Big(1-\mathrm{E}\big(e^{\mathcal{X}s}\big)\Big)^n \;= \\ &=\; -\sum_{n=1}^{\infty} \frac{1}{n}\left(-\sum_{m=1}^{\infty} \hat{\mu}_m \frac{s^m}{m!}\right)^n \;= \\ &=\; \hat{\mu}_1 s \,+\, \big(\hat{\mu}_2 - \hat{\mu}_1^2\big)\frac{s^2}{2!} \,+\, \big(\hat{\mu}_3 - 3\hat{\mu}_2\hat{\mu}_1 + 2\hat{\mu}_1^3\big)\frac{s^3}{3!} \,+\, \ldots \end{aligned} \tag{2.27}$$

The cumulants κ_n are obtained from the cumulant generating function through the n-th differentiation of $k(s)$ and calculating the derivative at $s = 0$:

$$\begin{aligned} \kappa_1 \;&=\; \frac{\partial k(s)}{\partial s}\Big|_{s=0} \;=\; \hat{\mu}_1 \;=\; \mu\,, \\ \kappa_2 \;&=\; \frac{\partial^2 k(s)}{\partial s^2}\Big|_{s=0} \;=\; \hat{\mu}_2 - \mu^2 \;=\; \sigma^2\,, \\ \kappa_3 \;&=\; \frac{\partial^3 k(s)}{\partial s^3}\Big|_{s=0} \;=\; \hat{\mu}_3 - 3\hat{\mu}_2\mu + 2\mu^3 \;=\; \mu_3\,, \\ &\;\;\vdots \\ \kappa_n \;&=\; \frac{\partial^n k(s)}{\partial s^n}\Big|_{s=0}\,, \\ &\;\;\vdots \end{aligned} \tag{2.16'}$$

As shown in equation (2.16) the first three cumulants coincide with the centered moments μ_1, μ_2, and μ_3. All higher cumulants are polynomials of two or more centered moments.

In probability theory the Laplace transform

$$\hat{f}(s) \;=\; \int_0^{\infty} e^{-sx}\, f_{\mathcal{X}}(x)\,\mathrm{d}x \;=\; \mathcal{L}\Big(f_{\mathcal{X}}(x)\Big)(s) \tag{2.28}$$

can be visualized as an expectation value that is closely related to the moment generating function: $\mathcal{L}\Big(f_{\mathcal{X}}(x)\Big)(s) = \mathrm{E}\big(e^{-s\mathcal{X}}\big)$ where $f_{\mathcal{X}}(x)$ is the probability density. The cumulative distribution function $F_{\mathcal{X}}(x)$ can be recovered by means of the inverse Laplace transform:

$$F_{\mathcal{X}}(x) = \mathcal{L}_s^{-1}\left(\frac{\mathrm{E}\big(e^{-s\mathcal{X}}\big)}{s}\right)(x) = \mathcal{L}_s^{-1}\left(\frac{\mathcal{L}\Big(f_{\mathcal{X}}(x)\Big)(s)}{s}\right)(x)\ .$$

We shall not use the Laplace transform here as a pendant to the moment generating function but we shall apply it in section 4.3.3.2 to the solution of chemical master equations where the inversion of the Laplace transform is discussed as well.

2.2.3 Characteristic functions

Like the moment generating function the *characteristic function* $\phi(s)$ of a random variable $\mathcal{X}$ completely describes the cumulative probability distribution $F(x)$. It is defined by

$$\phi(s) \;=\; \int_{-\infty}^{+\infty} \exp\left(\imath\, s\, x\right) \mathrm{d}F(x) \;=\; \int_{-\infty}^{+\infty} \exp\left(\imath\, s\, x\right) f(x)\, \mathrm{d}x\ , \tag{2.29}$$

where the integral over $\mathrm{d}F(x)$ is of Riemann-Stieltjes type. In case a probability density $f(x)$ exists for the random variable $\mathcal{X}$ the characteristic function is (almost) the Fourier transform of the density:[9]

$$\mathcal{F}\big(f(x)\big) \;=\; \tilde{f}(k) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x)\, e^{\imath\, k\, x}\, \mathrm{d}x\ . \tag{2.30}$$

From equation (2.29) follows the useful expression for the expansion in the discrete case

$$\phi(s) \;=\; \mathrm{E}\big(e^{\imath s \mathcal{X}}\big) \;=\; \sum_{n=-\infty}^{\infty} P_n\, e^{\imath n s}\ , \tag{2.29'}$$

that we shall use, for example, in solving the master equations for stochastic processes (chapters 3 and 4).

The characteristic function exists for all random variables since it is an integral of a bounded continuous function over a space of finite measure. There is a bijection between distribution functions and characteristic functions:

$$\phi_{\mathcal{X}}(s) \;=\; \phi_{\mathcal{Y}}(s) \iff F_{\mathcal{X}}(x) \;=\; F_{\mathcal{Y}}(x)\ .$$

---

[9] The difference between the Fourier transform $\tilde{f}(k)$ and the characteristic function $\phi(s)$ of a function $f(x)$,

$$\tilde{f}(k) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(x)\, \exp\left(+\imath\, k\, x\right) \mathrm{d}x \quad \text{and} \quad \phi(s) \;=\; \int_{-\infty}^{\infty} f(x) \exp\left(\imath\, s\, x\right) \mathrm{d}x\ ,$$

is only a matter of the factor $(\sqrt{2\pi})^{-1}$. The Fourier convention above is the one used in modern physics, for other convention see, e.g., *Mathematica* and section 4.4.2.

If a random variable $\mathcal{X}$ has moments up to k-th order, then the characteristic function $\phi(x)$ is k times continuously differentiable on the entire real line and vice versa if a characteristic function $\phi(x)$ has a k-th derivative at zero, then the random variable $\mathcal{X}$ has all moments up to k if k is even and up to $k-1$ if k is odd:

$$\mathrm{E}(\mathcal{X}^k) \;=\; (-\dot{\imath})^k \frac{\mathrm{d}^k \phi(s)}{\mathrm{d}s^k}\bigg|_{s=0} \quad \text{and} \quad \frac{\mathrm{d}^k \phi(s)}{\mathrm{d}s^k}\bigg|_{s=0} \;=\; \dot{\imath}^k \,\mathrm{E}(\mathcal{X}^k)\,. \tag{2.31}$$

An interesting example is presented by the Cauchy distribution (subsection 2.5.6) with $\phi(s) = \exp(|s|)$: It is not differentiable at $s = 0$ and the distribution has no moments including the expectation value.

The moment generating function is related to the probability generating function $g(s)$ (subsection 2.2.1) and the characteristic function $\phi(s)$ (subsection 2.2.3) by

$$g\left(e^s\right) \;=\; \mathrm{E}\left(e^{\mathcal{X}s}\right) \;=\; M_{\mathcal{X}}(s) \;\text{ and }\; \phi(s) \;=\; M_{\dot{\imath}\mathcal{X}}(s) \;=\; M_{\mathcal{X}}(\dot{\imath}s)\;.$$

All three generating functions are closely related as seen from the comparison of the expressions as expectation values

$$g(s) \;=\; \mathrm{E}(s^{\mathcal{X}})\,,\; M_{\mathcal{X}} \;=\; \mathrm{E}(e^{s\mathcal{X}})\,,\;\; \text{and} \;\; \phi(s) \;=\; \mathrm{E}(e^{\dot{\imath}s\mathcal{X}})\,,$$

but it may happen that not all three are existing. As said, characteristic functions exist for all probability distributions.

Table 2.2 Comparison of several common probability densities. Abbreviation and notations used in the table are: $\Gamma(r,x) = \int_x^\infty s^{r-1}e^{-s}\mathrm{d}s$ and
$\gamma(r,x) = \int_0^x s^{r-1}e^{-s}\mathrm{d}s$ are the upper and lower incomplete gamma function, respectively; $I_x(a,b) = B(x;a,b)/B(1;a,b)$ is the regularized incomplete beta function with $B(x;a,b) = \int_0^x s^{a-1}(1-s)^{b-1}\mathrm{d}s$. For more details see [116].

Name	Parameters	Support	pmf / pdf	cdf	Mean	Median	Mode	Variance	Skewness	Kurtosis	mgf	cf
Poisson $\pi(\alpha)$	$\alpha > 0 \in \mathbb{R}$	$k \in \mathbb{N}^0$	$\frac{\alpha^k}{k!}e^{-\alpha}$	$\frac{\Gamma(\lfloor k+1\rfloor,\alpha)}{\lfloor k\rfloor!}$	α	$\approx \lfloor \alpha+\frac{1}{3}-\frac{0.02}{\alpha}\rfloor$	$\lceil\alpha\rceil-1$	α	$\frac{1}{\sqrt{\alpha}}$	$\frac{1}{\alpha}$	$\exp\big(\alpha(e^s-1)\big)$	$\exp\big(\alpha(e^{\mathrm{i}s}-1)\big)$
Binomial $B(n,p)$	$n \in \mathbb{N}$ $p \in [0,1]$	$k \in \mathbb{N}^0$ $p \in [0,1]$	$\binom{n}{k}p^k(1-p)^{n-k}$	$I_{1-p} = (n-k, 1+k)$	np	$\lfloor np\rfloor$ or $\lceil np\rceil$	$\lfloor(n+1)p\rfloor$ or $\lfloor(n+1)p\rfloor-1$	$np(1-p)$	$\frac{1-2p}{\sqrt{np(1-p)}}$	$\frac{1-6p(1-p)}{np(1-p)}$	$(1-p+p^s)^n$	$(1-p+p^{\mathrm{i}s})^n$
Normal $\varphi(\nu,\sigma)$	$\nu \in \mathbb{R}$ $\sigma^2 \in \mathbb{R}^+$	$x \in \mathbb{R}$	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\nu)^2}{2\sigma^2}}$	$\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{x-\nu}{\sqrt{2\sigma^2}}\right)\right)$	ν	ν	ν	σ^2	0	0	$\exp(\nu s+\frac{1}{2}\sigma^2 s^2)$	$\exp(\mathrm{i}\nu s-\frac{1}{2}\sigma^2 s^2)$
chi-square $\chi^2(k)$	$k \in \mathbb{N}$	$x \in [0,\infty[$	$\frac{x^{\frac{k}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)}$	$\frac{\gamma(\frac{k}{2},\frac{x}{2})}{\Gamma(\frac{k}{2})}$	k	$\approx k\left(1-\frac{2}{9k}\right)^3$	$\max\{k-2,0\}$	$2k$	$\sqrt{\frac{8}{k}}$	$\frac{12}{k}$	$(1-2s)^{-\frac{k}{2}}$ for $s<\frac{1}{2}$	$(1-2\mathrm{i}s)^{-\frac{k}{2}}$
Logistic	$a \in \mathbb{R}, b>0$	$x \in \mathbb{R}$	$\frac{\mathrm{sech}^2\left((x-a)/2b\right)}{4b}$	$\frac{1}{1+\exp\left(-(x-a)/b\right)}$	a	a	a	$\pi^2b^2/3$	0	4.2	$\frac{\pi bs\,e^a s}{\sin(\pi bs)}$	$\frac{\mathrm{i}\pi bs\,e^a s}{\sin(\mathrm{i}\pi bs)}$
Laplace $b>0$	$\nu \in \mathbb{R}$	$x \in \mathbb{R}$	$\frac{1}{2b}e^{-\frac{\lvert x-\nu\rvert}{b}}$	$\begin{cases}\frac{1}{2}e^{-\frac{\nu-x}{b}}, & x<a\\ 1-\frac{1}{2}e^{-\frac{x-\nu}{b}}, & x\geq a\end{cases}$	ν	ν	ν	$2b^2$	0	3	$\frac{\exp(\nu s)}{1-b^2s^2}$ for $\lvert s\rvert<\frac{1}{b}$	$\frac{\exp(\mathrm{i}\nu s)}{1-b^2s^2}$
Uniform $a,b \in \mathbb{R}$	$a<b$	$x \in [a,b]$	$\begin{cases}\frac{1}{b-a}, & x\in[a,b]\\ 0 \text{ otherwise}\end{cases}$	$\begin{cases}0, & x<a\\ \frac{x-a}{b-a}, & x\in[a,b]\\ 1, & x\geq b\end{cases}$	$\frac{a+b}{2}$	$\frac{a+b}{2}$	$\widetilde{m}\in[a,b]$	$\frac{(b-a)^2}{12}$	0	$-\frac{6}{5}$	$\frac{e^{bs}-e^{as}}{(b-a)s}$	$\frac{e^{\mathrm{i}bs}-e^{\mathrm{i}as}}{\mathrm{i}(b-a)s}$
Cauchy $\gamma \in \mathbb{R}^+$	$x_0 \in \mathbb{R}$	$x \in \mathbb{R}$	$\frac{1}{\pi\gamma\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)}$	$\frac{1}{\pi}\arctan\left(\frac{x-x_0}{\gamma}\right)$	–	x_0	x_0	–	–	–	–	$\exp(\mathrm{i}x_0 s-\gamma\lvert s\rvert)$

2.3 Common probability distributions

Before entering a discussion of individual probability distributions we present an overview over the important characteristics of the most frequently used distributions in table 2.2. Poisson, binomial and normal distributions and transformations in limits between them are discussed in this section. The central limit theorem and the law of large numbers are presented in a separate section following the normal distribution. We have also listed several less common but nevertheless frequently used probability distributions, which are of importance for special purposes. In the forthcoming chapters 3, 4, and 5 dealing with stochastic processes and applications we shall make use of them.

2.3.1 The Poisson distribution

The Poisson distribution, named after the French physicist and mathematician Siméon Denis Poisson, is a discrete probability distribution expressing the probability of occurrence of independent events within a given interval. A popular example is dealing with the arrivals of phone calls, e-mail, an other independent events within a fixed time interval Δt. The expected number of occurring events per time interval, α, is the only parameter of the distribution $\pi_k(\alpha)$, which returns the probability that k events are recorded during Δt. In physics and chemistry the Poisson process is the stochastic basis of first order reactions, for example radioactive decay or irreversible first order chemical reactions, and the Poisson distribution is the probability distribution underlying the time course of particle numbers, atoms or molecules, fulfilling

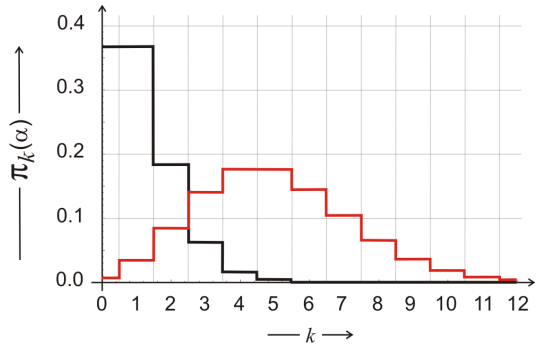


Fig. 2.6 The Poisson probability density. Two examples of Poisson distributions, $\pi_k(\alpha) = \alpha^k e^{-\alpha}/k!$, with $\alpha = 1$ (black) and $\alpha = 5$ (red) are shown. The distribution with the larger α has the mode shifted further to the right and a thicker tail.

$\mathrm{d}N(t) = -\alpha\, N(t)\,\mathrm{d}t$. The events to be counted need not be on the time axis, the interval can also be defined as a given distance, area, or volume.

Despite its major importance in physics and biology the Poisson distribution with the probability mass function (pmf) $\pi_k(\alpha)$, is a fairly simple mathematical object. As said it contains a single parameter only, the real valued positive number α:

$$P(\mathcal{X} = k) \;=\; \pi_{\,k}(\alpha) \;=\; \frac{e^{-\alpha}}{k!}\,\alpha^k \;;\quad k \in \mathbb{N}^0 \;, \tag{2.32}$$

where $\mathcal{P} = \mathcal{X}$ is a random variable with Poissonian density. As an exercise we leave to verify the following properties:[10]

$$\sum_{k=0}^{\infty} \pi_{\,k} = 1\,,\;\; \mu \;=\; \sum_{k=0}^{\infty} k\,\pi_{\,k} = \alpha \;\;\text{ and }\;\; \hat{\mu}_2 \;=\; \sum_{k=0}^{\infty} k^2\,\pi_{\,k} = \alpha + \alpha^2$$

Examples of Poisson distributions with two different parameter values, $\alpha = 1$ and 5, are shown in figure 2.6. The cumulative distribution function (cdf) is obtained by summation

$$P(\mathcal{X} \le k) \;=\; \exp(-\alpha) \sum_{j=0}^{\lfloor k \rfloor} \frac{\alpha^j}{j!} \;=\; \frac{\Gamma(\lfloor k+1 \rfloor, \alpha)}{\lfloor k \rfloor !} \;, \tag{2.33}$$

where $\Gamma(x, y)$ is the incomplete Gamma function.

By means of a Taylor series expansion[11] we can find the generating function of the Poisson distribution,

$$g(s) \;=\; e^{\alpha(s-1)} \;. \tag{2.34}$$

From the generating function we calculate easily

$$g'(s) \;=\; \alpha\, e^{\alpha(s-1)} \;\;\text{ and }\;\; g''(s) \;=\; \alpha^2\, e^{\alpha(s-1)} \;.$$

Expectation value and second moment follow straightforwardly from the exercise above or from equation(2.25):

---

[10] In order to be able to solve the problems some basic infinite series should be recalled: $e = \sum_{n=0}^{\infty} \frac{1}{n!}$, $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ for $|x| < \infty$, $e = \lim_{n\to\infty}(1 + \frac{1}{n})^n$, and $e^{-\alpha} = \lim_{n\to\infty}(1 - \frac{\alpha}{n})^n$.

[11] The Taylor series is named after the English mathematician Brook Taylor who invented the calculus of finite differences in 1715. Earlier series expansions of the MacLaurin series type in the 17th century are attributed to the Scottish mathematician James Gregory. The MacLaurin series is a Taylor expansion around the origin and was named after Colin MacLaurin.

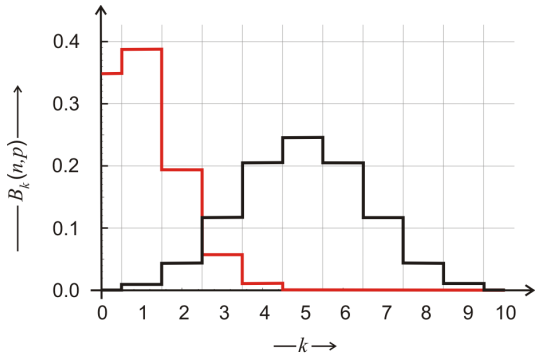


Fig. 2.7 The binomial probability density. Two examples of binomial distributions, $B_k(n,p) = \binom{n}{k} p^k (1-p)^{n-k}$, with $n = 10$, $p = 0.5$ (black) and $p = 0.1$ (red) are shown. The former distribution is symmetric with respect to the expectation value $\mathrm{E}(B_k) = n/2$, and accordingly has zero skewness. The latter case is asymmetric with positive skewness (see figure 2.3).

$$\mathrm{E}(\mathcal{X}) \;=\; g'(1) \;=\; \alpha \;, \tag{2.34a}$$

$$\mathrm{E}(\mathcal{X}^2) \;=\; g'(1) + g''(1) \;=\; \alpha + \alpha^2 \;, \quad \text{and} \tag{2.34b}$$

$$\mathrm{var}(\mathcal{X}) \;=\; \alpha \;. \tag{2.34c}$$

Both, the expectation value and the variance are equal to the parameter α and hence, the standard deviation amounts to $\sigma(\mathcal{X}) = \sqrt{\alpha}$. Accordingly, the Poisson distribution is the discrete prototype of a distribution fulfilling a $\sqrt{N}$-law. This remarkable property of the Poisson distribution is not limited to the second moment: The *factorial moments*, $\langle \mathcal{X}^r \rangle_f$, fulfil the equation

$$\langle \mathcal{X}^r \rangle_f \;=\; \mathrm{E}\Big(\mathcal{X}(\mathcal{X}-1)\ldots(\mathcal{X}-r+1)\Big) \;=\; \alpha^r \;, \tag{2.34d}$$

which is easily verified by direct calculation.

2.3.2 The binomial distribution

The binomial distribution, $B(n,p)$, expresses the cumulative result of n independent trials with two-valued outcomes, for example, yes-no decisions or successive coin tosses as we have discussed already in sections 1.2 and 1.5:

$$\mathcal{S}_n \;=\; \sum_{i=1}^{n} \mathcal{X}_i \,, i \in \mathbb{N}_{>0};\, n \in \mathbb{N}_{>0} \;. \tag{1.23'}$$

In general, we assume that head is obtained with probability p and tail with probability $q = 1 - p$. The $\mathcal{X}_i$'s are called Bernoulli random variables named after the Swiss mathematician Jakob Bernoulli, and the sequence of events, $\mathcal{S}_n$, is named Bernoulli process after him (section 3.1.3.1). The corresponding random variable is said to have a Bernoulli or binomial distribution::

$$\begin{aligned} P(\mathcal{S}_n = k) \;&=\; B_k(n,p) \;=\; \binom{n}{k} p^k \, q^{n-k} \,, \\ q = 1-p \;\;&\text{and}\;\; k \in \mathbb{N};\, k \leq n \;. \end{aligned} \tag{2.35}$$

Two examples are shown in figure 2.7. The distribution with $p = 0.5$ is symmetric with respect to $k = n/2$.

The generating function for the single trial is $g(s) = q + p\,s$. Since we have n independent trials the complete generating function is

$$g(s) \;=\; (q \,+\, p\,s)^n \;=\; \sum_{k=0}^{n} \binom{n}{k} q^{n-k}\, p^k \, s^k \;. \tag{2.36}$$

From the derivatives of the generating function,

$$g'(s) \;=\; n\,p\,(q + p\,s)^{n-1} \;\;\text{and}\;\; g''(s) \;=\; n(n-1)\,p^2\,(q + p\,s)^{n-2} \;,$$

we compute readily expectation value and variance:

$$\mathrm{E}(\mathcal{S}_n) \;=\; g'(1) \;=\; n\,p \;, \tag{2.36a}$$

$$\mathrm{E}(\mathcal{S}_n^2) \;=\; g'(1) \,+\, g''(1) \;=\; np \,+\, n^2p^2 \,-\, np^2 \;=\; n\,p\,q \,+\, n^2\,p^2 \;, \tag{2.36b}$$

$$\mathrm{var}(\mathcal{S}_n) \;=\; n\,p\,q \;, \;\; \text{and} \tag{2.36c}$$

$$\sigma(\mathcal{S}_n) \;=\; \sqrt{npq} \;. \tag{2.36d}$$

For $p = 1/2$, the case of the unbiased coin, we are dealing with the *symmetric binomial distribution* with $\mathrm{E}(\mathcal{S}_n) = n/2$, $\mathrm{var}(\mathcal{S}_n) = n/4$, and $\sigma(\mathcal{S}_n) = \sqrt{n}/2$. We note that the expectation value is proportional to the number of trials, n, and the standard deviation is proportional to its square root, $\sqrt{n}$.

Relation between binomial and Poisson distribution. The binomial distribution $B(n,p)$ can be transformed into a Poisson distribution $\pi(\alpha)$ in the limit $n \to \infty$. In order to show this we start from

$$B_k(n,p) \;=\; \binom{n}{k} p^k \,(1-p)^{n-k} \,;\; k \in \mathbb{N}, k \leq n \;.$$

The symmetry parameter p is assumed to vary with n, $p\,(n) \,=\, \alpha/n$ for $n \in \mathbb{N}{>}\,0$, and thus we have

$$B_k\left(n,\frac{\alpha}{n}\right) \;=\; \binom{n}{k}\left(\frac{\alpha}{n}\right)^k\left(1-\frac{\alpha}{n}\right)^{n-k}\;,\;(k\in\mathbb{N},k\le n)\;.$$

We let n go to infinity for fixed k and start with $B_0(n,p)$:

$$\lim_{n\to\infty} B_0\left(n,\frac{\alpha}{n}\right) \;=\; \lim_{n\to\infty}\left(1-\frac{\alpha}{n}\right)^n \;=\; e^{-\alpha}\;.$$

Now we compute the ratio of two consecutive terms, B_{k+1}/B_k:

$$\frac{B_{k+1}\left(n,\frac{\alpha}{n}\right)}{B_k\left(n,\frac{\alpha}{n}\right)} \;=\; \frac{n-k}{k+1}\cdot\left(\frac{\alpha}{n}\right)\cdot\left(1-\frac{\alpha}{n}\right)^{-1} \;=\; \frac{\alpha}{k+1}\cdot\left[\left(\frac{n-k}{n}\right)\cdot\left(1-\frac{\alpha}{n}\right)^{-1}\right]\;.$$

Both terms in the square brackets converge to one as $n\to\infty$, and hence we find:

$$\lim_{n\to\infty}\frac{B_{k+1}\left(n,\frac{\alpha}{n}\right)}{B_k\left(n,\frac{\alpha}{n}\right)} \;=\; \frac{\alpha}{k+1}\;.$$

From the two results we compute all terms starting from the limit of B_0,

$$\begin{aligned}
\lim_{n\to\infty} B_0 &= \exp(-\alpha) \;\text{ and find}\\
\lim_{n\to\infty} B_1 &= \alpha\exp(-\alpha)\;,\\
\lim_{n\to\infty} B_2 &= \alpha^2\exp(-\alpha)/2!\;,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
\lim_{n\to\infty} B_k &= \alpha^k\exp(-\alpha)/k!\;. \qquad\qquad \square
\end{aligned}$$

Accordingly we have verified Poisson's limit law:

$$\lim_{n\to\infty} B_k\left(n,\frac{\alpha}{n}\right) \;=\; \pi_k(\alpha)\,,\;k\in\mathbb{N}\;. \tag{2.37}$$

It is worth keeping in mind that the limit was performed in a peculiar way since the symmetry parameter $p(n)=\alpha/n$ was shrinking with increasing n and as a matter of fact vanished in the limit of $n\to\infty$.

2.3.3 The normal distribution

The normal or Gaussian distribution is of central importance in probability theory because many distributions converge to it in the limit of large numbers since the central limit theorem (CLT) states that under mild conditions the sum of a large number of random variables is approximately normal distributed (section 2.4.2). The normal distribution is a *stable distribution* (section 3.2.5) and this fact is not unrelated to the central limit theorem.

The normal distribution has several advantageous technical features. It is the only absolutely continuous distribution, which has only zero cumulants except the first two corresponding to expectation value and variance, which have the straightforward meaning of the position and the width of the distribution. For given variance the normal distribution has the largest informational entropy of all distributions on $\Omega = \mathbb{R}$ (section 2.1.3). As a matter of fact, the mean μ does not enter the expression for the entropy of the normal distribution (table 2.1),

$$H(\sigma) \;=\; \frac{1}{2}\left(1 \,+\, \log\left(2\pi\sigma^2\right)\right) \;, \tag{2.21'}$$

or in other words, shifting the normal distribution along the x-axis does not change the entropy of the distribution.

The normal distribution is basic for the estimate of statistical errors and thus we shall discuss it in some detail. Because of this fact, the normal distribution is extremely popular in statistics and sometimes 'overapplied'. Many empirical values are not symmetrically distributed but skewed to the right but nevertheless they are often analyzed by means of normal distributions. The log-normal distribution [286] or the Pareto distribution, for example, might do better in such cases. Statistics based on normal distribution is not robust in the presence of outliers where a description by more heavy-tailed distributions like Student's t-distribution is superior. Whether or not the tails have more weight in the distribution can be easily checked by means of the excess kurtosis: Student's distribution has an excess kurtosis of

$$\gamma_2 \;=\; \begin{cases} \frac{6}{\nu-4} & \text{for } \nu > 4 \;, \\ \infty & \text{for } 2 < \nu \leq 4 \;, \quad \text{and} \\ \text{undefined} & \text{otherwise} \,, \end{cases}$$

which is always positive, whereas the excess kurtosis of the normal distribution is zero.

The density of the normal distribution[12] is

$$f_{\mathcal{N}}(x) \;=\; \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{with} \quad \int_{-\infty}^{+\infty} f(x)\,dx \;=\; 1 \;, \tag{2.38}$$

and the corresponding random variable $\mathcal{X}$ has the moments $\mathrm{E}(\mathcal{X}) = \mu$, $\mathrm{var}(\mathcal{X}) = \sigma^2$, and $\sigma(\mathcal{X}) = \sigma$. For many purposes it is convenient to use the normal density in centered and normalized form ($\sigma^2 = 1$), which is called the standard normal distribution or the Gaussian bell-shaped curve:[12]

[12] The notations applied here for the normal distribution are: $\mathcal{N}(\mu, \sigma)$ in general, and $F_{\mathcal{N}}(x; \mu, \sigma)$ for the cumulative distribution or $f_{\mathcal{N}}(x; \mu, \sigma)$ for the density. Commonly, the parameters, (μ, σ) are omitted when no misinterpretation is possible. For standard stable distributions (section 2.5.7) a variance $\gamma^2 = \sigma^2/2$ is applied.

$$f_{\mathcal{N}}(x;0,1) \;=\; \varphi(x) \;=\; \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2} \;\text{ with }\; \int_{-\infty}^{+\infty}\varphi(x)\,\mathrm{d}x \;=\; 1\;, \qquad (2.38')$$

In this form we have $\mathrm{E}(\tilde{\mathcal{X}}) = 0$, $\mathrm{var}(\tilde{\mathcal{X}}) = 1$, and $\sigma(\tilde{\mathcal{X}}) = 1$.

Integration of the density yields the distribution function

$$P(\mathcal{X} \le x) \;=\; F(x) \;=\; \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,\mathrm{d}u \;=\; F_{\mathcal{N}}(x)\;. \qquad (2.39)$$

The function $F_{\mathcal{N}}(x)$ is not available in analytical form, but it can be easily formulated in terms of a special function, the error function, $\mathrm{erf}(x)$. This function as well as its complement, $\mathrm{erfc}(x)$, are defined by

$$\mathrm{erf}(x) \;=\; \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}\,\mathrm{d}u \;\text{ and }\; \mathrm{erfc}(x) \;=\; \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-u^2}\,\mathrm{d}u\;,$$

and are available in tables and in standard mathematical packages.[13] Examples of the normal density $f_{\mathcal{N}}(x)$ and the integrated distribution $F_{\mathcal{N}}(x)$ with different values of the standard deviation σ were shown in figure 1.21. The normal distribution is also used in statistics to define confidence intervals: 68.2 % of the data points lie within an interval $\mu \pm \sigma$, 95.4 % within an interval $\mu \pm 2\sigma$, and eventually 99,7 % with an interval $\mu \pm 3\sigma$.

A Poisson density with sufficiently large values of α resembles a normal density (figure 2.8) and it can be shown indeed that the two curves become more and more similar with increasing α:[14]

$$\pi_k(\alpha) \;=\; \frac{\alpha^k}{k!}\,e^{-\alpha} \;\approx\; \frac{1}{\sqrt{2\pi\,\alpha}}\,\exp\!\left(-\frac{(k-\alpha)^2}{2\alpha}\right) \;\text{ for }\; \alpha \gg 1\;. \qquad (2.40)$$

We present a short proof based on the moment generating functions for the approximation of the standardized Poisson distribution by a standard normal distribution. The Poisson variable $\mathcal{X}_\alpha$ with $P(\mathcal{X}_\alpha = k) = \pi_k(\alpha)$ is standardized to $\mathcal{Y}_\alpha = (\mathcal{X}_\alpha - \alpha)/\sqrt{\alpha}$ and we obtain for the moment generating functions:

$$M_{\mathcal{X}_\alpha}(s) = \mathrm{E}\left(e^{\mathcal{X}_\alpha s}\right) = \exp\bigl(\alpha(e^s - 1)\bigr) \;\;\Longrightarrow\;\; M_{\mathcal{Y}_\alpha}(s) = \mathrm{E}\left(\exp\Bigl(\frac{\mathcal{X}_\alpha - \alpha}{\sqrt{\alpha}}s\Bigr)\right).$$

The next steps are taking the limit $n \to \infty$, expansion of the exponential function, and truncation after the first non-vanishing term [276]:

---

[13] We remark that $\mathrm{erf}(x)$ and $\mathrm{erfc}(x)$ are not normalized in the same way as the normal density is: $\mathrm{erf}(x) + \mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_0^\infty \exp(-u^2)\mathrm{d}u = 1$, but $\int_0^\infty \varphi(u)\mathrm{d}u = \frac{1}{2}\int_{-\infty}^{+\infty}\varphi(u)\mathrm{d}u = \frac{1}{2}$.

[14] It is important to memorize that k is a discrete variable on the l.h.s. whereas it is continuous on the r.h.s. of equation (2.40).

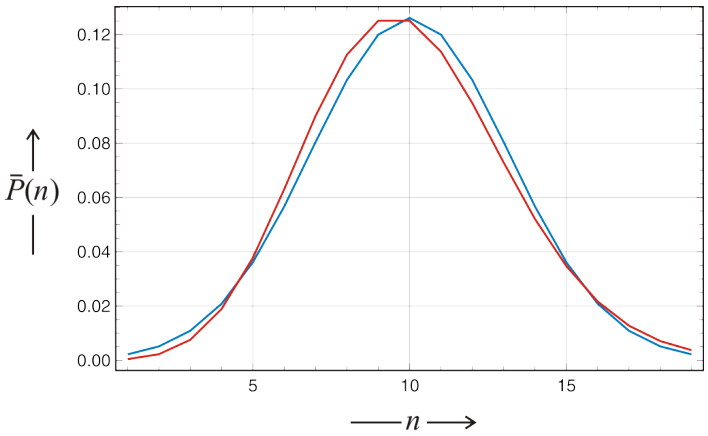


Fig. 2.8 Comparison between Poisson and normal density. The figure compares the Poisson density with parameter α (red) and a best fit normal distribution with mean $\mu = \alpha$ and standard deviation $\sigma = \sqrt{\alpha}$ (blue) according to equation (2.40). Parameter choice: $\alpha = 10$.

$$\lim_{\alpha\to\infty} M_{y_\alpha}(s) = \lim_{\alpha\to\infty} \mathrm{E}\left(\exp\Big(\frac{\mathcal{X}_\alpha - \alpha}{\sqrt{\alpha}}s\Big)\right) = \lim_{\alpha\to\infty} e^{-\sqrt{\alpha}s}\,\mathrm{E}\left(\exp\Big(\frac{\mathcal{X}_\alpha s}{\sqrt{\alpha}}\Big)\right) =$$
$$= \lim_{\alpha\to\infty} e^{-\sqrt{\alpha}\,s}\exp\big(\alpha(e^{s/\sqrt{\alpha}} - 1)\big) =$$
$$= \lim_{\alpha\to\infty} \exp\big(s^2/2 + s^3/(6\sqrt{\alpha}) + \ldots\big) = \exp(s^2/2) \qquad \square$$

In the limit of large α we obtain indeed the moment generating function of the standardized normal distribution, $\mathcal{N}(0,1)$. The result is an example of the central limit theorem that will be presented and analyzed in section 2.4.2. We shall require this approximation of the Poissonian distribution by a normal distribution in section 3.4.5 for the derivation of a chemical Langevin equation.

The normal density function $f_\mathcal{N}(x)$ has, among other remarkable properties, derivatives of all orders. Each derivative can be written as product of $f_\mathcal{N}(x)$ by a polynomial, of the order of the derivative, known as Hermite polynomial. The function $f_\mathcal{N}(x)$ decreases to zero very rapidly as $|x| \to \infty$. The existence of all derivatives makes the *bell-shaped* Gaussian curve $x \to f(x)$ particularly smooth, and the moment generating function of the normal distribution is especially attractive (see subsection 2.2.2) since $M(s)$ can be obtained directly by integration:

$$\begin{aligned} M(s) &= \int_{-\infty}^{+\infty} e^{x\,s}\, f(x)\,\mathrm{d}x \;=\; \int_{-\infty}^{+\infty} \exp\left(x\,s \,-\, \frac{x^2}{2}\right) \mathrm{d}x \;= \\ &= \int_{-\infty}^{+\infty} e^{(\frac{s^2}{2} \,-\, \frac{(x-s)^2}{2})}\,\mathrm{d}x \;=\; e^{s^2/2} \int_{-\infty}^{+\infty} f(x-s)\,\mathrm{d}x \;= \\ &= e^{s^2/2}\ . \end{aligned} \tag{2.41}$$

All raw moments of the normal distribution are defined by the integrals

$$\hat{\mu}_n \;=\; \int_{-\infty}^{+\infty} x^n\, f(x)\, dx\ . \tag{2.42}$$

They can be obtained, for example, by successive differentiation of $M(s)$ with respect to s (subsection 2.2.2). In order to obtain the moments more efficiently we expand the first and the last expression in equation (2.41) in a power series of s,

$$\begin{aligned} &\int_{-\infty}^{+\infty} \left(1 \,+\, x\,s \,+\, \frac{(x\,s)^2}{2!} \,+\ldots+\, \frac{(x\,s)^n}{n!} \,+\ldots\right) f(x)\,\mathrm{d}x \;= \\ &=\quad 1 \,+\, \frac{s^2}{2} \,+\, \frac{1}{2!}\left(\frac{s^2}{2}\right)^2 \,+\ldots+\, \frac{1}{n!}\left(s^2 2\right)^n \,+\ldots\ , \end{aligned}$$

or express it in terms of the moments $\hat{\mu}_n$,

$$\sum_{n=0}^{\infty} \frac{\hat{\mu}_n}{n!}\, s^n \;=\; \sum_{n=0}^{\infty} \frac{1}{2^n\, n!}\, s^{2n}\ ,$$

from which we compute the moments of $\varphi(x)$ by putting equal the coefficients of the powers in s on both sides of the expansion, and find for $n \geq 1$:[15]

$$\hat{\mu}_{2n-1} \;=\; 0 \;\text{ and }\; \hat{\mu}_{2n} \;=\; \frac{(2n)!}{2^n\, n!}\ . \tag{2.43}$$

All odd moments vanish because of symmetry. In case of the fourth moment, kurtosis, a kind of standardization is common, which assigns zero excess kurtosis, $\gamma_2 = 0$ to the normal distribution. In other words, excess kurtosis monitors peak shape with respect to the normal distribution: Positive excess kurtosis implies peaks that are sharper than the normal density, negative excess kurtosis peaks that are broader than the normal density (figure 2.3).

---

[15] The definite integrals are:

$$\textstyle\int_{-\infty}^{+\infty} x^n \exp\left(-x^2\right)\mathrm{d}x = \begin{cases} \sqrt{\pi} & n = 0 \\ 0 & n \geq 1;\ \text{odd} \\ \frac{(n-1)!!}{2^{n/2}}\sqrt{\pi} & n \geq 2;\ \text{even} \end{cases}\ ,$$

where $(n-1)!! = 1 \cdot 3 \cdot \ldots \cdot (n-1)$ is the *double factorial*.

As already said all cumulants (2.16) of the normal distribution except $\kappa_1 = \mu$ and $\kappa_2 = \sigma^2$ are zero, since the moment generating function of the general normal distribution with mean μ and variance σ^2 is of the form

$$M_{\mathcal{N}}(s) \;=\; \exp\Big(\mu s \,+\, \frac{1}{2}\,\sigma^2\,s^2\Big)\;. \tag{2.44}$$

The expression for the standardized Gaussian distribution is the special case with $\mu = 0$ and $\sigma^2 = 1$. Eventually, we list also the characteristic function of the general normal distribution

$$\phi_{\mathcal{N}}(s) \;=\; \exp\Big(\dot{\imath}\mu s \,-\, \frac{1}{2}\,\sigma^2\,s^2\Big)\;, \tag{2.45}$$

which will be used, for example, in the derivation of the central limit theorem (section 2.4.2).

2.3.4 Multivariate normal distributions

In applications to real world problems in science it is often necessary to consider probability distributions in multiple dimensions. Then, a random vector, $\vec{\mathcal{X}} = (\mathcal{X}_1, \ldots, \mathcal{X}_n)$ with the joint probability distribution

$$P(\mathcal{X}_1 = x_1, \ldots, \mathcal{X}_n = x_n) \;=\; p\,(x_1, \ldots, x_n) \;=\; p\,(\mathbf{x})\;.$$

replaces the random variable $\mathcal{X}$. This multivariate normal probability density can be written as

$$f(\mathbf{x}) \;=\; \frac{1}{\sqrt{(2\pi)^n\,|\boldsymbol{\Sigma}|}}\;\exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{t}}\;\boldsymbol{\Sigma}^{-1}\,(\mathbf{x}-\boldsymbol{\mu})\Big)\;.$$

The vector $\boldsymbol{\mu}$ consists of the (raw) first moments along the different coordinates, $\boldsymbol{\mu}(\mu_1, \ldots, \mu_n)$ and the variance-covariance matrix $\boldsymbol{\Sigma}$ contains the n variances in the diagonal and the covariances are combined as off-diagonal elements

$$\boldsymbol{\Sigma} \;=\; \begin{pmatrix} \mathrm{var}(\mathcal{X}_1) & \mathrm{cov}(\mathcal{X}_1,\mathcal{X}_2) & \ldots & \mathrm{cov}(\mathcal{X}_1,\mathcal{X}_n) \\ \mathrm{cov}(\mathcal{X}_2,\mathcal{X}_1) & \mathrm{var}(\mathcal{X}_2) & \ldots & \mathrm{cov}(\mathcal{X}_2,\mathcal{X}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(\mathcal{X}_n,\mathcal{X}_1) & \mathrm{cov}(\mathcal{X}_n,\mathcal{X}_2) & \ldots & \mathrm{var}(\mathcal{X}_n) \end{pmatrix} \;=\; \begin{pmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{12} & \sigma_{22} & \ldots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \ldots & \sigma_{nn} \end{pmatrix}$$

which is symmetric, $\mathrm{cov}(\mathcal{X}_i,\mathcal{X}_j) = \mathrm{cov}(\mathcal{X}_j,\mathcal{X}_i) = \sigma_{ij}$, by the definition of covariances.

Mean and variance are given by $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}$ and the variance-covariance matrix Σ, the moment generating function expressed in the dummy vector variable $\mathbf{s} = (s_1, \ldots, s_n)$ is of the form

$$M(\mathbf{s}) \;=\; \exp\left(\boldsymbol{\mu}^{\mathrm{t}}\mathbf{s}\right)\cdot\exp\left(\frac{1}{2}\,\mathbf{s}^{\mathrm{t}}\,\Sigma\mathbf{s}\right)\;,$$

and, finally, the characteristic function is given by

$$\phi(\mathbf{s}) \;=\; \exp\left(\mathrm{i}\,\boldsymbol{\mu}^{\mathrm{t}}\mathbf{s}\right)\cdot\exp\left(-\frac{1}{2}\,\mathbf{s}^{\mathrm{t}}\,\Sigma\mathbf{s}\right)$$

Without showing the details we remark that this particulary simple characteristic function implies that all moments higher than order two can be expressed in terms of first and second moments, in particular expectation values, variances, and covariances. To give an example that we shall require later in subsection 3.4.2, the fourth order moments can be derived from

$$\begin{aligned}
\mathrm{E}(\mathcal{X}_i^4) &= 3\,\sigma_{ii}^2\;,\\
\mathrm{E}(\mathcal{X}_i^3\mathcal{X}_j) &= 3\,\sigma_{ii}\sigma_{ij}\;,\\
\mathrm{E}(\mathcal{X}_i^2\mathcal{X}_j^2) &= \sigma_{ii}\sigma_{jj}+2\,\sigma_{ij}^2\;,\\
\mathrm{E}(\mathcal{X}_i^2\mathcal{X}_j\mathcal{X}_k) &= \sigma_{ii}\sigma_{jk}+2\,\sigma_{ij}\sigma_{ik}\quad\text{and}\\
\mathrm{E}(\mathcal{X}_i\mathcal{X}_j\mathcal{X}_k\mathcal{X}_l) &= \sigma_{ij}\sigma_{kl}+\sigma_{li}\sigma_{jk}+\sigma_{ik}\sigma_{jl}\;,
\end{aligned} \tag{2.46}$$

with $i,j,k,l\in\{1,2,3,4\}$.

The entropy of the multivariate normal distribution is readily calculated and appears as a straightforward extension of equation (2.21) to higher dimensions:

$$\begin{aligned}
H(f) &= -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(\mathbf{x})\,\ln f(\mathbf{x})\,\mathrm{d}\mathbf{x} \;=\\
&= \frac{1}{2}\Big(n+\ln\big((2\pi)^n|\boldsymbol{\Sigma}|\big)\Big)\;,
\end{aligned} \tag{2.47}$$

where $|\boldsymbol{\Sigma}|$ is the determinant of the variance-covariance matrix.

The multivariate normal distribution presents and excellent example for discussing the difference between uncorrelatedness and independence. Two random variables are independent if

$$f_{\mathcal{XY}}(x,y) \;=\; f_{\mathcal{X}}(x)\cdot f_{\mathcal{Y}}(y)\;\;\forall\,x,y\;,$$

whereas uncorrelatedness of two random variables requires

$$\begin{aligned}
\sigma_{\mathcal{XY}} \;=\; \mathrm{cov}(\mathcal{X},\mathcal{Y}) \;&=\; 0 \;=\; \mathrm{E}(\mathcal{XY})-\mathrm{E}(\mathcal{X})\mathrm{E}(\mathcal{Y})\quad\text{or}\\
\mathrm{E}(\mathcal{XY}) \;&=\; \mathrm{E}(\mathcal{X})\mathrm{E}(\mathcal{Y})\;,
\end{aligned}$$

which is only factorizability of the joint expectation value. The covariance between two independent random variables vanishes and hence:

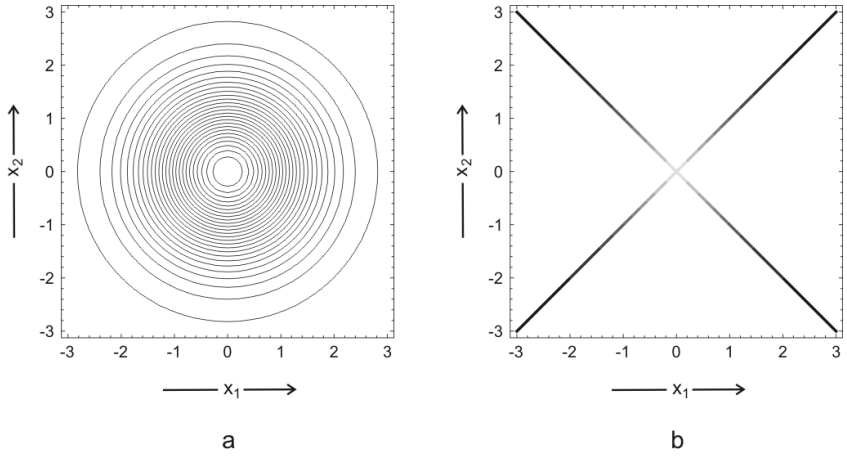


Fig. 2.9 Uncorrelated but not independent normal distributions. The figure compares two different joint densities, which have identical marginal densities. The contour plot on the l.h.s. (**a**) shows the joint distribution $f(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2+x_2^2)}$, the contour lines are circles equidistant in f and plotted for $f = 0.03, 0.09, \ldots, 0.153$. The marginal distributions of this joint distribution are standard normal distributions in x_1 or x_2. The density in **b** is derived from one random variable $\mathcal{X}_1$ with standard normal density $f(x_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2}$ and a second random variable that is modulated by a perfect coin flip: $\mathcal{X}_2 = \mathcal{X}_1 \cdot \mathcal{W}$ with $\mathcal{W} = \pm 1$. The two variables $\mathcal{X}_1$ and $\mathcal{X}_2$ are uncorrelated but not independent.

$$
\begin{aligned}
\mathrm{E}(\mathcal{XY}) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy\, f_{\mathcal{X},\mathcal{Y}}(x,y)\,\mathrm{d}x\mathrm{d}y \;= \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy\, f_{\mathcal{X}}(x) f_{\mathcal{Y}}(y)\,\mathrm{d}x\mathrm{d}y \;= \\
&= \int_{-\infty}^{+\infty} x\, f_{\mathcal{X}}(x)\mathrm{d}x \int_{-\infty}^{+\infty} y\, f_{\mathcal{Y}}(y)\mathrm{d}y \;= \mathrm{E}(\mathcal{X})\mathrm{E}(\mathcal{Y})\ . \qquad \square
\end{aligned}
$$

We remark that the proof made nowhere use of the fact that the variables are normally distributed and the statement *independent variables are uncorrelated* holds in full generality. The inverse, however, is not true as has been shown by means of specific examples [323]: Uncorrelated random variables $\mathcal{X}_1$ and $\mathcal{X}_2$, which both have the same (marginal) normal distribution, need not be independent. The construction of such a contradicting example starts from a two dimensional random vector $\vec{\mathcal{X}} = (\mathcal{X}_1, \mathcal{X}_2)^{\mathrm{t}}$, which obeys a bivariate normal distribution with mean $\boldsymbol{\mu} = (0,0)^{\mathrm{t}}$, variance $\sigma_1^2 = \sigma_2^2 = 1$ and covariance $\mathrm{cov}(\mathcal{X}_1, \mathcal{X}_2) = 0$

$$f(x_1,x_2) \;=\; \frac{1}{2\pi}\,\exp\left(-\frac{1}{2}\,(x_1,x_2)\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}\right) \;=$$
$$=\; \frac{1}{2\pi}\,e^{-\frac{1}{2}\,(x_1^2+x_2^2)} \;=\; \frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}\,x_1^2}\cdot\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}\,x_2^2} \;=\; f(x_1)\cdot f(x_2)\,.$$

The two random variables are independent. Next we introduce a modification in one of the two random variables: $\mathcal{X}_1$ remains unchanged and has the density $f(x_1) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x_1^2\right)$, whereas the second random variable is modulated by an ideal coin flip, $\mathcal{W}$ with the density $f(w) = \frac{1}{2}\big(\delta(w+1)+\delta(w-1)\big)$. In other words, we have $\mathcal{X}_2 = \mathcal{W}\cdot\mathcal{X}_1 = \pm\mathcal{X}_1$ with equal weights for both signs, and accordingly the density function is $f(x_2) = \frac{1}{2}f(x_1) + \frac{1}{2}f(-x_1) = f(x_1)$, since the normal distribution with zero mean $\mathrm{E}(\mathcal{X}_1) = 0$ is symmetric, $f(x_1) = f(-x_1)$. Equality of the two distribution functions with the same normal distribution can also be derived directly:

$$\begin{aligned}P(\mathcal{X}_2\le x) \;&=\; \mathrm{E}\big(P(\mathcal{X}_2\le x|\mathcal{W})\big) \;=\\ &=\; P(\mathcal{X}_1\le x)P(\mathcal{W}=1) \;+\; P(-\mathcal{X}_1\le x)P(\mathcal{W}=-1) \;=\\ &=\; F_{\mathcal{N}}(x)\frac{1}{2} \;+\; F_{\mathcal{N}}(x)\frac{1}{2} \;=\; F_{\mathcal{N}}(x) \;=\; P(\mathcal{X}_1\le x)\;.\end{aligned}$$

The covariance of $\mathcal{X}_1$ and $\mathcal{X}_2$ is readily calculated,

$$\begin{aligned}\mathrm{cov}(\mathcal{X}_1\mathcal{X}_2) \;&=\; \mathrm{E}(\mathcal{X}_1\mathcal{X}_2) \;-\mathrm{E}(\mathcal{X}_1)\cdot\mathrm{E}(\mathcal{X}_2) \;=\; \mathrm{E}(\mathcal{X}_1\mathcal{X}_2) \;-\; 0 \;=\\ &=\; \mathrm{E}\Big(\mathrm{E}(\mathcal{X}_1\mathcal{X}_2)|\mathcal{W}\Big) \;=\; \mathrm{E}(\mathcal{X}_1^2)P(\mathcal{W}=1)+\mathrm{E}(-\mathcal{X}_1^2)P(\mathcal{W}=-1) \;=\\ &=\; 1\,\frac{1}{2} \;+\; (-1)\,\frac{1}{2} \;=\; 0\;,\end{aligned}$$

and hence $\mathcal{X}_1$ and $\mathcal{X}_2$ are uncorrelated. The two random variables, however, are not independent because

$$\begin{aligned}p\,(x_1,x_2) \;&=\; P(\mathcal{X}_1=x_1,\mathcal{X}_2=x_2) \;=\\ &=\; \frac{1}{2}P(\mathcal{X}_1=x_1,\mathcal{X}_2=x_1)+\frac{1}{2}P(\mathcal{X}_1=x_1,\mathcal{X}_2=-x_1)\Big) \;=\\ &=\; \frac{1}{2}p\,(x_1)+\frac{1}{2}p\,(x_1) \;=\; p\,(x_1)\;,\\ f(x_1,x_2) \;&=\; f(x_1) \;\neq\; f(x_1)\cdot f(x_2)\;,\end{aligned}$$

since $f(x_1) = f(x_2)$. Lack of independence follows also simply from $|\mathcal{X}_1| = |\mathcal{X}_2|$. The example is illustrated in figure 2.9: The fact that marginal distributions are identical does not imply that the joint distribution is also the same! The statement about independence, however, can be made stronger and then it turns out to be true: *"If random variables have a multivariate*

normal distribution and are pairwise uncorrelated, then the random variables are always independent." [323].

The marginal distributions of a multivariate normal distribution are obtained straightforwardly by simply dropping the marginalized variables. If $\vec{\mathcal{X}} = (\mathcal{X}_i, \mathcal{X}_j, \mathcal{X}_k)$ is a multivariate, normally distributed variable with the mean vector $\boldsymbol{\mu} = (\mu_i, \mu_j, \mu_k)$ and variance-covariance matrix $\boldsymbol{\Sigma}$, then after elimination of $\mathcal{X}_j$ the marginal joint distribution of the vector $\tilde{\mathcal{X}} = (\mathcal{X}_i, \mathcal{X}_k)$ is multivariate normal with the mean vector $\tilde{\boldsymbol{\mu}} = (\mu_i, \mu_k)$ and the variance-covariance matrix

$$\tilde{\boldsymbol{\Sigma}} \;=\; \begin{pmatrix} \Sigma_{ii} & \Sigma_{ik} \\ \Sigma_{ki} & \Sigma_{kk} \end{pmatrix} \;=\; \begin{pmatrix} \mathrm{var}(\mathcal{X}_i) & \mathrm{cov}(\mathcal{X}_i, \mathcal{X}_k) \\ \mathrm{cov}(\mathcal{X}_k, \mathcal{X}_i) & \mathrm{var}(\mathcal{X}_k) \end{pmatrix} \;.$$

It is worth noticing that non-normal bivariate distributions have been constructed, which have normal marginal distributions [263].

2.4 Regularities at large numbers

Having seen in previous sections 1.9.1 and 2.3.3 that probability distribution may converge to the normal distribution in the limit of large numbers of trials, one is inclined to conjecture a more general regularity behind these special cases. The prototype of a system that can be extrapolated to large numbers is the sequence of Bernoulli trials, the probabilistic outcome of which is given by the binomial distribution and we shall show that indeed the binomial distribution converges asymptotically to the normal distribution. The generalization to a sequence of independent variables with distributions different from binomial and an average according to equation (1.23')

$$\overline{\mathcal{X}} \;=\; \frac{1}{n}\,\mathcal{S}_n \;=\; \frac{1}{n}\,(\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n) \;,$$

no matter whether they are identically distributed or not, is the subject of the central limit theorem (CLT). The sufficient condition is finite expectation value and finite variance for all random variables $\mathcal{X}_j$. The CLT concerns the distribution of the random variable $\mathcal{S}_n$ in the limit $n \to \infty$. Two other regularities concern first and second moment of $\mathcal{S}_n$: The *law of large numbers* guarantees convergence of the sum $\mathcal{S}_n$ to the expectation value in strong and weak forms

$$\lim_{n\to\infty} \mathcal{S}_n \;=\; n\,\mu \;,$$

and the *law of the iterated logarithm* confines the fluctuations:

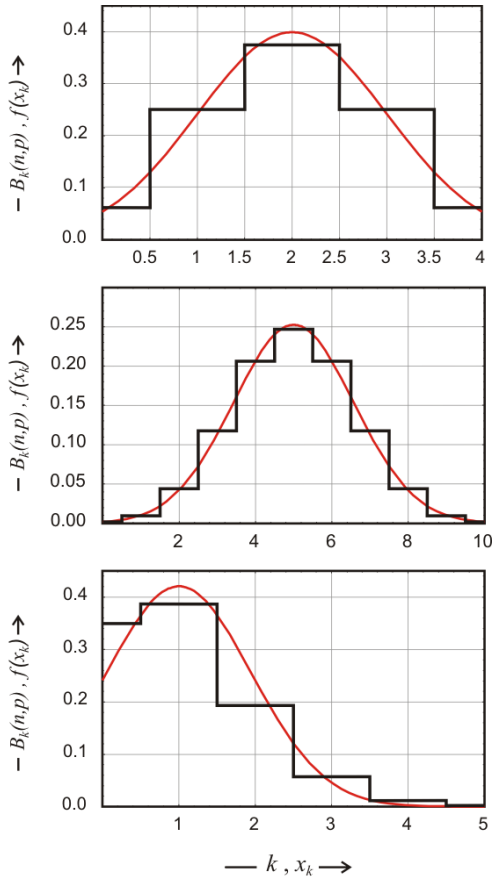


Fig. 2.10 A fit of the normal distribution to the binomial distribution. The curves represent normal densities (red), which were fit to the points of the binomial distribution (black). The three examples. Parameter choice for the binomial distribution: $(n = 4, p = 0.5)$, $(n = 10, p = 0.5)$, and $(n = 5, p = 0.1)$, for the upper, middle, and lower plot, respectively.

$$\limsup_{n\to\infty} (\mathcal{S}_n - n\,\mu) \;=\; +\sigma\,\sqrt{n}\,\sqrt{2\ln(\ln n)} \quad \text{and}$$
$$\liminf_{n\to\infty} (\mathcal{S}_n - n\,\mu) \;=\; -\sigma\,\sqrt{n}\,\sqrt{2\ln(\ln n)}\;.$$

For larger values of n the iterated logarithm, $\ln(\ln n)$, is a very slowly increasing function of n and hence the upper and lower bound to the stochastic variable are not too different from $\sqrt{n}$ (figure 2.12). The law of the iterated logarithm is the rigorous final answer to the conjectures $\sqrt{n}$-law that we have mentioned already several times.

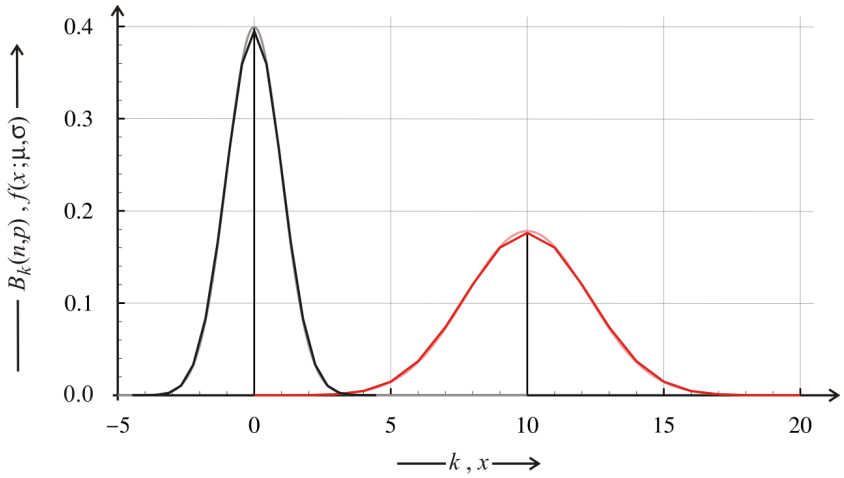


Fig. 2.11 Standardization of the binomial distribution. The figure shows a symmetric binomial distribution $B(20, \frac{1}{2})$, which is centered around $\mu = 10$ (red). The transformation yields a binomial distribution centered around the origin with unit variance: $\sigma = \sigma^2 = 1$ (black). The pink curve is a normal density $f_{\mathcal{N}}(x) = \exp\big(-(x-\mu)^2/(2\sigma^2)\big)/\sqrt{2\pi\sigma^2}$ with the parameters $\mu = 10$ and $\sigma^2 = np\,(1-p) = 5$, and the gray line is a standardized normal density $\varphi(x)$ ($\mu = 0, \sigma^2 = 1$), respectively.

2.4.1 From binomial to normal distributions

The expression *normal distribution* actually originated from the fact that many distributions can be transformed in a natural way for large numbers n to yield the distribution $F_{\mathcal{N}}(x)$. Here we derive this result for the binomial distribution where it appears most natural. A binomial density,

$$B_k(n,p) \;=\; \binom{n}{k} p^k\,(1-p)^{n-k}\,,\;\; 0 \le k \le n\;,$$

becomes a normal density through extrapolation to large values of n at constant p.[16] The transformation from the binomial distribution to the normal distribution is properly done in two steps (see also [69, pp.210-217]): (i) standardization and (ii) taking the limit $n \to \infty$.

At first we make the binomial distribution comparable to the standard normal density, $\varphi(x) = e^{-x^2/2}/\sqrt{2\pi}$, by shifting the maximum towards $x = 0$ and adjusting the width (figures 2.10 and 2.11). For $0 < p < 1$ and $q = 1 - p$ the

[16] This is different from an extrapolation performed in a previous section2.3.2 because the limit $\lim_{n\to\infty} B_k(n, \alpha/n) = \pi_k(\alpha)$ leading to the Poisson distribution was performed for vanishing $p = \alpha/n$.

discrete variable k is replaced by a new variable ξ :[17]

$$\xi \;=\; \frac{k \,-\, np}{\sqrt{npq}} \;;\;\; 0 \le k \le n \;.$$

Instead of the variables $\mathcal{X}_k$ and $\mathcal{S}_k$ in equation (1.23') new random variables, $\mathcal{X}_k^*$ and $\mathcal{S}_n^* = \sum_{k=1}^n \mathcal{X}_k^*$ are introduced, which account for centering around $x = 0$ and adjustment to the width of a standard Gaussian, $\varphi(x)$, by making use of the expectation value, $\mathrm{E}(\mathcal{S}_n) = np$, and the standard deviation, $\sigma(\mathcal{S}_n) = \sqrt{npq}$, of the binomial distribution.

The theorem of de Moivre and Laplace states now that for k in a neighborhood of $k = np - |\xi| \le c$ with c being an arbitrary fixed positive constant – the approximation

$$\binom{n}{k} p^k q^{n-k} \;\approx\; \frac{1}{\sqrt{2\pi npq}}\, e^{-\xi^2/2} \,;\; p+q=1,\, p>0,\, ,q>0 \qquad (2.48')$$

becomes exact in the sense that the ratio of the lhs to the rhs converges to one as $n \to \infty$ [133, section VII.3]. The convergence is uniform with respect to k in the range specified above. In order to proof the convergence we transform the lhs by making use of Stirling's formula, $n! \approx n^n e^{-n}\sqrt{2\pi n}$ as $n \to \infty$:

$$\binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k} \approx \sqrt{\frac{n}{2\pi k(n-k)}} \left(\Big(\frac{k}{np}\Big)^{-k} \Big(\frac{n-k}{nq}\Big)^{-(n-k)} \right) .$$

Next we introduce the variable ξ and transform to the exponential function

$$\begin{aligned} \binom{n}{k} p^k q^{n-k} \;&\approx\; \frac{1}{\sqrt{2\pi npq}} \left(\Big(1+\xi\sqrt{\frac{q}{np}}\Big)^{-k} \Big(1-\xi\sqrt{\frac{p}{nq}}\Big)^{-(n-k)} \right) \;= \\ &=\; \frac{1}{\sqrt{2\pi npq}} \;\; e^{\ln\left(\left(1+\xi\sqrt{\frac{q}{np}}\right)^{-k} \left(1-\xi\sqrt{\frac{p}{nq}}\right)^{-(n-k)} \right)} . \end{aligned}$$

Then expansion of the logarithm yields

$$\begin{aligned} &\ln\left(\Big(1+\xi\sqrt{\frac{q}{np}}\Big)^{-k} \Big(1-\xi\sqrt{\frac{p}{nq}}\Big)^{-(n-k)} \right) \;= \\ &=\; -k\,\ln\left(1+\xi\sqrt{\frac{q}{np}}\right) \;-\; (n-k)\,\ln\left(1-\xi\sqrt{\frac{p}{nq}}\right) \;. \end{aligned}$$

Next making use of the series expansion $\ln(1 \pm \gamma) \approx \pm\gamma - \gamma^2/2 \pm \gamma^3/3 - \ldots$, truncation after the second term, and inserting $k = np + \xi\sqrt{npq}$ and $n - k = nq - \xi\sqrt{npq}$ we find

[17] The new variable ξ depends on k and n, but for short we dispense from subscripts.

$$\ln\left(\left(1+\xi\sqrt{\frac{q}{np}}\right)^{-k}\left(1-\xi\sqrt{\frac{p}{nq}}\right)^{-(n-k)}\right) \;=$$

$$= \;-\,(np+\xi\sqrt{npq}\,)\,\left(\xi\sqrt{\frac{q}{np}}-\xi^2\frac{q}{np}+\ldots\right)-$$

$$-\;(nq-\xi\sqrt{npq}\,)\,\left(-\xi\sqrt{\frac{p}{nq}}-\xi^2\frac{p}{nq}+\ldots\right).$$

Evaluation of the expressions eventually yields

$$\ln\left(\left(1+\xi\sqrt{\frac{q}{np}}\right)^{-k}\left(1-\xi\sqrt{\frac{p}{nq}}\right)^{-(n-k)}\right) \;\approx\; -\frac{\xi^2}{2}$$

and thereby we have proved the conjecture (2.48'). □

A comparison of figures 2.10 and 2.11 shows that the convergence of the binomial distribution to the normal distribution is particularly effective in the symmetric case, $p = q = 0.5$. The difference is somewhat larger for $p = 0.1$. A value of $n = 20$ is sufficient to make the difference hardly recognizable with the unaided eye. Figure 2.11 shows the effect of standardization on the binomial distribution.

In the context of the central limit theorem (section 2.4.2) it is appropriate to formulate the theorem of de Moivre and Laplace in a slightly different way: The distribution of the standardized random variable $\mathcal{S}_n^*$ with a binomial distribution converges in the limit of large numbers n to the normal distribution $\varphi(x)$ on any finite constant interval $]\,a,b]$ with $a < b$:

$$\lim_{n\to\infty} P\left(\left(\frac{\mathcal{S}_n-np}{\sqrt{npq}}\right)\in\,]a,b]\right) \;=\; \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}\,\mathrm{d}x\;. \tag{2.48}$$

In the proof [69, p. 215-217] the definite integral $\int_a^b \varphi(x)\,\mathrm{d}x$ is partitioned into n small segments like in Riemannian integration, where the segments still reflect the discrete distribution. In the limit $n \to \infty$ the partition becomes finer and eventually converges to the continuous function described by the integral. In the sense in section 1.8.1 we are dealing with convergence to a limit in distribution.

2.4.2 Central limit theorem

In addition to the transformation of the binomial distribution into the normal distribution analyzed in the previous section 2.4.1 we have already encountered two cases where other probability distributions approached the normal distribution in the limit of large numbers n: (i) the distribution of scores for rolling n dice simultaneously (section 1.9.1) and (ii) the Poisson distribution

(section 2.3.3). Therefore it is obvious to conjecture a more general role of the normal distribution in the limit of large numbers. The Russian mathematician Aleksandr Lyapunov pioneered the formulation and derivation of the generalization known as *central limit theorem* (CLT) [298, 299]. Research on CLT was continued since then and has been completed at least for practical purposes through extensive studies during the twentieth century [5, 401].

The *central limit theorem* comes in various stronger and weaker forms. We mention here three of them:

(i) The so-called *classical central limit theorem* is commonly associated with the names of the Finnish mathematician Jarl Waldemar Lindeberg [289] and the French mathematician Paul Pierre Lévy [280]. It is the most common version used in practice. In essence, the Lindeberg-Lévy central limit theorem is noting but the generalization of the de Moivre-Laplace theorem (2.48) that was used in the previous section 2.4.1 to proof the transition from the binomial to the normal distribution in the limit $n \to \infty$.

The generalization proceeds from Bernoulli variables to *independent and identically distributed* (iid) random variables $\mathcal{X}_i$. The distribution is arbitrary, need not be specified and the only requirements are a finite expectation value and a finite variance: $\mathrm{E}(\mathcal{X}_i) = \mu < \infty$ and $\mathrm{var}(\mathcal{X}_i) = \sigma^2 < \infty$. Again we consider the sum of n random variables, $\mathcal{S}_n = \sum_{i=1}^{n} \mathcal{X}_i$, standardize to yield $\mathcal{X}_i^*$ and $\mathcal{S}_n^*$, and instead of equation (2.48) we obtain

$$\lim_{n\to\infty} P\left(\frac{\mathcal{S}_n - n\,\mu}{\sqrt{n}\,\sigma} \in]a,b]\right) \;=\; \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\,\mathrm{d}x. \tag{2.49}$$

For every segment $a < b$ the arbitrary initial distribution converges to the normal distribution in the limit $n \to \infty$. Although this is already a remarkable extension of the validity in the limit of the normal distribution, the results can be made more general.

(ii) Lyapunov's earlier version of the central limit theorem [298, 299] requires only independent and not necessarily identically distributed variables $\mathcal{X}_i$ with finite expectation values, μ_i, and variances, σ_i^2 provided a criterium called *Lyapunov condition* is fulfilled by the sum of variances $s_n^2 = \sum_{i=1}^{n} \sigma_1^2$,

$$\lim_{n\to\infty} \frac{1}{s^{2+\delta}} \sum_{i=1}^{n} \mathrm{E}\left(|\mathcal{X}_i - \mu_i|^{2+\delta}\right) \;=\; 0\;. \tag{2.50}$$

Then the sum $\sum_{i=1}^{n}(\mathcal{X}_i - \mu_i)/s_n$ converges in distribution in the limit $n \to \infty$ to the standard normal distribution:

$$\frac{1}{s_n} \sum_{i=1}^{n} (\mathcal{X}_i - \mu_i) \;\xrightarrow{d}\; \mathcal{N}(0,1)\;. \tag{2.51}$$

Whether or not a given sequence of random variables fulfils the Lyapunov condition is commonly verified in practice by setting $\delta = 1$.

(iii) Lindeberg showed in the year 1922 [290] that a weaker condition than Lyapunov's condition is sufficient to guarantee the convergence in distribution to the standard normal distribution:

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} \mathrm{E}\Big((X_i - \mu_i)^2 \cdot \mathbf{1}_{|\mathcal{X}_i - \mu_i| > \epsilon s_n} \Big) \;=\; 0 \;, \tag{2.52}$$

where $\mathbf{1}_{|\mathcal{X}_i - \mu_i| > \epsilon s_n}$ is the indicator function (1.27a) identifying sample space

$$\{|\mathcal{X}_i - \mu_i| > \epsilon\, s_n\} \;\doteq\; \{\omega \in \Omega :\, |\mathcal{X}_i(\omega) - \mu_i| > \epsilon\, s_n\} \;.$$

If a sequence of random variables satisfies Lyapunov's condition it satisfies also Lindeberg's condition but the converse does not hold in general. Lindeberg's condition is sufficient but not necessary in general, and the condition for necessity is

$$\max_{i=1,\ldots,n} \frac{\sigma_i^2}{s_n^2} \;\to\; 0 \quad \text{as} \quad n \to \infty \;,$$

or, in other words, the Lindeberg condition is fulfilled if and only if the central limit theorem holds.

The three versions of the central limit theorem are related to each other: Lindeberg's condition (iii) is the most general form and hence both the classical CLT (i) and the Lyapunov CLT (ii) can be derived as special cases from (iii). It is worth noticing, however, that (i) does not follow necessarily from (ii), because (i) requires a finite second moment whereas the condition for (ii) is a finite moment of order $(2 + \delta)$.

In summary the central limit theorem for a sequence of independent random variables $\mathcal{S}_n = \sum_{i=1}^{n} \mathcal{X}_i$ with finite means, $\mathrm{E}(\mathcal{X}_i) = \mu_i < \infty$, and variances, $\mathrm{var}(\mathcal{X}_i) = \sigma_i^2 < \infty$, states that the sum $\mathcal{S}_n$ converges in distribution to a standardized normal random variable $\mathcal{N}(0, 1)$ without any further restriction on the densities of the variables. The literature on the central limit theorem is enormous and several proofs with many variants have been derived (see, for example, [68] or [69, pp. 222-224]). We dispense here from a repetition of this elegant proof that makes use of the characteristic function, and present only the key equation for the convergence where the number n approaches infinity with s being fixed:

$$\lim_{n\to\infty} \mathrm{E}\Big(e^{\mathrm{i}s\,\mathcal{S}_n^*} \Big) \;=\; \lim_{n\to\infty} \left(1 \,-\, \frac{s^2}{2n} \left(1 + \varepsilon\Big(\frac{s}{\sqrt{n}} \Big) \right) \right)^n \;=\; e^{-s^2/2} \;, \tag{2.53}$$

with ε being any small positive constant.

For practical applications used in the *statistics of large samples* the central limit theorem as encapsulated in equation (2.53) is turned into a rough approximation

$$P(\sigma\sqrt{n}\, x_1 \,<\, \mathcal{S}_n - n\,\mu \,<\, \sigma\sqrt{n}\, x_2) \;\approx\; F_{\mathcal{N}}(x_2) \,-\, F_{\mathcal{N}}(x_1) \;. \tag{2.54}$$

The spread around the mean μ is obtained by setting $x = x_1 = -x_2$

$$P(|\mathcal{S}_n - n\,\mu| < \sigma\sqrt{n}\,x) \;\approx\; 2F_{\mathcal{N}}(x) \;-\; 1\;. \tag{2.54'}$$

In *pre-computer time* equation (2.54) has been used extensively with the aid of tabulations of the functions $F_{\mathcal{N}}(x)$ and $F_{\mathcal{N}}^{-1}(x)$, which are still found in most textbooks of statistics .

2.4.3 Law of large numbers

The *law of large numbers* states that in the limit of infinitely large samples the sum of random variable converges to the expectation value:

$$\frac{1}{n}\mathcal{S}_n \;=\; \frac{1}{n}\,(\mathcal{X}_1 + \mathcal{X}_2 + \ldots + \mathcal{X}_n) \;\;\to\;\; \mu \quad \text{for} \quad n \to \infty\;.$$

In its strong form the law can be expressed as

$$P\left(\lim_{n\to\infty}\frac{1}{n}\mathcal{S}_n \;=\; \mu\right) \;=\; 1\;. \tag{2.55a}$$

In other words, the sample average converges almost certainly to the expectation value.

The weaker form of the law of large numbers is written as

$$P\left(\lim_{n\to\infty}\left|\frac{1}{n}\mathcal{S}_n - \mu\right| \;>\; \varepsilon\right) \;=\; 0\;, \tag{2.55b}$$

and implies convergence in probability: $\mathcal{S}_n/n \xrightarrow{P} \mu$. The weak law states that for any sufficiently large sample there exists a zone $\mu \pm \varepsilon$ around the expectation value – no matter how small it is – such that the average of the observed quantity will come so close to the expectation value that it lies within the zone.

It is illustrative to visualize the difference between the strong and the weak law also from a dynamical perspective: The weak law says that the average $\mathcal{S}_n/n$ will be near μ provided n is sufficiently large. The sample, however, may rarely but infinitely often leave the zone and fulfil $|\mathcal{S}_n/n - \mu| > \varepsilon$, and the frequency with which this happens is of measure zero. The strong law asserts that such excursions will almost certainly never happen and the inequality $|\mathcal{S}_n/n - \mu| < \varepsilon$ holds for all n that are large enough.

The law of large numbers can be derived as a straightforward consequence of the central limit theorem (2.49) [69, pp.227-233]. For any fixed but arbitrary constant $\varepsilon > 0$ we have

$$\lim_{n\to\infty} P\left(\left|\frac{\mathcal{S}_n}{n} - \mu\right| < \varepsilon\right) = 1 \,. \tag{2.56}$$

The constant ε is fixed and therefore we can define a positive constant ℓ that fulfils $\ell < \varepsilon\sqrt{n}/\sigma$ and for which

$$\left\{\left|\frac{\mathcal{S}_n - n\mu}{\sqrt{n}\sigma}\right| < \ell\right\} \implies \left\{\left|\frac{\mathcal{S}_n - n\mu}{n}\right| < \varepsilon\right\} \,,$$

and hence,

$$P\left(\left|\frac{\mathcal{S}_n - n\mu}{\sqrt{n}\sigma}\right| < \ell\right) \leq P\left(\left|\frac{\mathcal{S}_n - n\mu}{n}\right| < \varepsilon\right) \,,$$

provided n is sufficiently large. Now we go back to equation (2.49), and choose a symmetric interval $a = -\ell$ and $b = +\ell$ for the integral. Then the lhs of the inequality converges to $\int_{-l}^{+l} \exp(-x^2/2)\mathrm{d}x/\sqrt{2\pi}$ in the limit $n \to \infty$. For any $\delta > 0$ we can choose ℓ so large that the value of the integral exceeds $1 - \delta$ and we get for sufficiently large values of n

$$P\left(\left|\frac{\mathcal{S}_n}{n} - \mu\right| < \varepsilon\right) = 1 - \delta \,. \tag{2.57}$$

This result proves that the law of large numbers (2.56) is a corollary of (2.49). □

Related to and a consequence of equation (2.56) is Chebyshev's inequality for random variables $\mathcal{X}$ that have a finite second moment, which is named after the Russian mathematician Pafnuty Lvovich Chebyshev :

$$P(|\mathcal{X}| \geq c) \leq \frac{\mathrm{E}(\mathcal{X}^2)}{c^2} \tag{2.58}$$

and which is true for any constant $c > 0$. We dispense here from a proof that is found in [69, pp. 228-233]. By means of Chebyshev's inequality the law of large numbers (2.56) can extended to a sequence of independent random variables $\mathcal{X}_j$ with different expectation values and variances, $\mathrm{E}(\mathcal{X}_j) = \mu^{(j)}$ and $\mathrm{var}(\mathcal{X}_j) = \sigma_j^2$, with the restriction that there exists a constant $\Sigma^2 < \infty$ such that $\sigma_j^2 \leq \Sigma^2$ is fulfilled for all $\mathcal{X}_j$. Then we have for each $c > 0$:

$$\lim_{n\to\infty} P\left(\left|\frac{\mathcal{X}_1 + \ldots + \mathcal{X}_n}{n} - \frac{\mu^{(1)} + \ldots + \mu^{(n)}}{n}\right| < c\right) = 1 \,. \tag{2.59}$$

The main message of the law of large numbers is that for a sufficiently large numbers of independent events the statistical errors in the sum will vanish and the mean converges to the exact expectation value. Hence, the law of large numbers provides the basis for the assumption of convergence in mathematical statistics (section 2.6).

2.4.4 Law of the iterated logarithm

The *law of the iterated logarithm* consists of two asymptotic regularities derived for sums of random variables, which are related to the central limit theorem and the law of large numbers, and in a way complete the predictions of both. The name of the law points at the appearance of the function 'log log' in the forthcoming expressions – it does not refer to the notion of the *iterated logarithm* in computer science[18] – and the derivation is attributed to the two Russian scholars of mathematics Aleksandr Khinchin [248] and Andrey Kolmogorov [257]. The proof in the degree of generality used here has been provided later [130, 198]. The law of the iterated logarithm provides upper and lower bounds for the values of sums of random variables and in this ways confines the size of fluctuations.

For a sum of n independent and identically distributed (iid) random variables with expectation value $\mathrm{E}(\mathcal{X}_i) = \mu$ and finite variance $\mathrm{var}(\mathcal{X}) = \sigma^2 < \infty$,

$$\mathcal{S}_n \;=\; \mathcal{X}_1 + \mathcal{X}_1 + \ldots + \mathcal{X}_n \;,$$

the following two limits are fulfilled with probability one:

$$\limsup_{n\to\infty} \frac{\mathcal{S}_n - n\,\mu}{\sqrt{2\,n\,\ln(\ln\,n)}} \;=\; +\,|\sigma| \quad \text{and} \tag{2.60a}$$

$$\liminf_{n\to\infty} \frac{\mathcal{S}_n - n\,\mu}{\sqrt{2\,n\,\ln(\ln\,n)}} \;=\; -\,|\sigma| \;. \tag{2.60b}$$

The two theorems (2.60) are equivalent and this follows directly from the symmetry of the standardized normal distribution, $\mathcal{N}(0,1)$. We dispense here from the presentation of a proof for the law of the iterated logarithm that can be found, for example, in a monograph by Henry McKean [314] or in a publication by William Feller [130]. For the purpose of illustration we compare with the already mentioned heuristic $\sqrt{n}$-law (see section 1.1), which is based on the properties of the symmetric standardized binomial distribution $B(n,p)$ with $p = \frac{1}{2}$. Accordingly we have $2\sigma/n = 1/\sqrt{n}$ and consequently most values of $\mathcal{S}_n - n\mu$ lie in the interval $-|\sigma| \le \mathcal{S}_n \le +|\sigma|$. The corresponding result from the law of the iterated logarithm is

---

[18] In computer science the iterated logarithm of n is commonly written $\log^* n$ and represents the number of times the logarithmic function must be iteratively applied before the result is is less than or equal to one:

$$\log^* \;\doteq\; \begin{cases} 0 & \text{if} \quad n \le 1 \;, \\ 1 \,+\, \log^*(\log n) & \text{if} \quad n > 1 \;. \end{cases}$$

The iterated logarithm is well defined for base 'e', for base '2' and in general for any base greater than $e^{1/e} = 1.444667\ldots$.

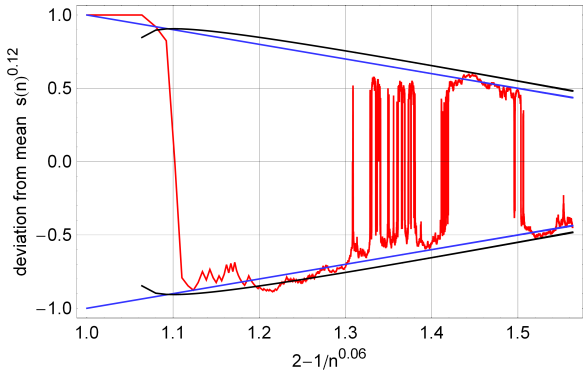


Fig. 2.12 Illustration of the law of the iterated logarithm. The picture shows the sum of the scorers of a sequence of Bernoulli trials with the outcome $\mathcal{X}_i = \pm 1$ and $\mathcal{S}_n = \sum_{i=1}^{n} \mathcal{X}_i$: The standardized sum, $\mathcal{S}(n)/n - \mu = s(n) - \mu = s(n)$ since $\mu = 0$, is shown as a function of n. In order to make the plot illustrative we adopt the scaling of the axes proposed by Dean Foster [148] that yields a straight line for the function $\sigma(n) = 1/\sqrt{n}$: On the x-axis we plot $x(n) = 2 - 1/n^{0.06}$, and this results in the following pairs of values $(x, n) = (1, 1)$, $(1.129, 10)$, $(1.241, 100)$, $(1.339, 1000)$, $(1.564, 10^6)$, $(1.810, 10^{12})$, and $(2, \infty)$. The y-axis is split into two halves corresponding to positive and negative values of $s(n)$. In the positive half we plot $s(n)^{0.12}$ and in the negative half $-|s(n)|^{0.12}$ in order to yield symmetry between the positive and the negative zone. The two blue curves provide an envelope $\mu \pm \sigma = \mu \pm \sqrt{1/n}$, and the two black curves present the results of the law of the iterated logarithm, $\mu \pm \sqrt{(2\ln(\ln n))/n}$. Note that the function $\ln(\ln n)$ adopts negative values for $1 < x < 1.05824$ $(1 < n < 2.71828)$.

$$-\sqrt{\frac{2\,\ln\,(\ln\,n)}{n}} \;\leq\; \mathcal{S}_n \;\leq\; +\sqrt{\frac{2\,\ln\,(\ln\,n)}{n}}$$

with probability one. One particular case of iterated Bernoulli trials – tosses of a fair coin – is shown in figure 2.12, where the envelope of the sum $\mathcal{S}_n$ of the cumulative score of n trials, $\pm\sqrt{2\,\ln(\ln n)/n}$ is compared with the results of the square root n law, $\mu \pm \sigma = \pm\sqrt{1/n}$. We remark that quite frequently the sum takes on values close to the envelopes. The special importance of the results of the law of the iterated logarithm for the Wiener process will be discussed later (section 3.2.2.2).

In essence, we may summarize the results of this section in three statements, which are part of *large sample theory*: For independent and identically distributed (iid) random variables $\mathcal{X}_i$ with $\mathcal{S}_n = \sum_{i=1}^{n} \mathcal{X}_i$ with $\mathrm{E}(\mathcal{X}_i) = \mathrm{E}(\mathcal{X}) = \mu$ and finite variance $\mathrm{var}(\mathcal{X}_i) = \sigma < \infty$ we have the three large sample results:

(i) the *law of large numbers*: $\mathcal{S}_n \to n\,\mathrm{E}(\mathcal{X}) = n\,\mu$,

(ii) the *law of the iterated logarithm*: $\begin{cases} \limsup \frac{(S_n - n\mu)}{\sqrt{2n \ln(\ln n)}} & \to \; +|\sigma| \\ \liminf \frac{(S_n - n\mu)}{\sqrt{2n \ln(\ln n)}} & \to \; -|\sigma| \end{cases}$, and

(iii) the *central limit theorem*: $\frac{1}{\sqrt{n}}\big(\mathcal{S}_n - n\,\mathrm{E}(\mathcal{X})\big) \to \mathcal{N}(0,1)\,.$

The theorem (i) defines the limit of the sample average, theorem (ii) determines the size of fluctuations and theorem (iii), eventually, refers to the limiting distribution function, which turns out to be the normal distribution. All three theorems can be extended in their range of validity to independent random variables with arbitrary distributions provided mean and variance are finite.

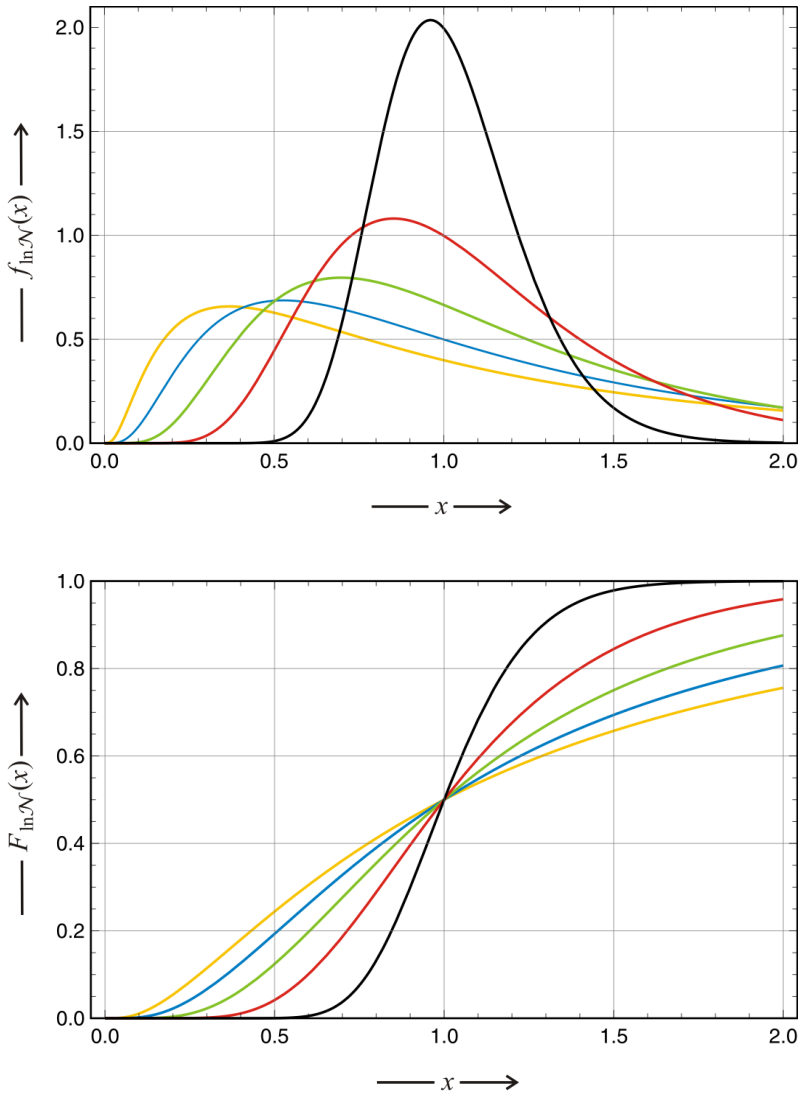


Fig. 2.13 The log-normal distribution. The log-normal distribution, $\ln\mathcal{N}(\mu,\sigma)$, is defined on the positive real axis, $x \in]0,\infty[$ and has the probability density (pdf)

$$f_{\ln\mathcal{N}}(x) = \exp\Big(-(\ln x-\mu)^2/(2\sigma^2)\Big) \Big/ \left(x\sqrt{2\pi\sigma^2}\right)$$

and the cumulative distribution function (cdf)

$$F_{\ln\mathcal{N}}(x) = \left(1+\operatorname{erf}\Big((\ln x-\mu)/\sqrt{2\sigma^2}\Big)\right) \Big/ \, 2 \; .$$

The two parameters are confined by the relations $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Parameter choice and color code: $\mu = 0$, σ =0.2 (black), 0.4 (red), 0.6 (green), 0.8 (blue), and 1.0 (yellow).

2.5 Further probability distributions

In the previous section 2.3 we presented the three most relevant probability distributions: (i) the Poisson distribution because it describes the distribution of occurrence of independent events, (ii) the binomial distribution dealing with independent trials with two outcomes, and (iii) the normal distribution being the limiting distribution of large numbers of individual events irrespectively of the statistics of single events. In this section we shall discuss seven more or less arbitrarily selected distributions, which play an important role in science and/or in statistics. The presentation here is inevitably rather brief and for reading of a detailed treatise we refer to [235, 236]. Other probability distributions will be mentioned together with the problems to which they are applied, for example the Erlang distribution in the discussion of the Poisson process (section 3.2.2.4) and the Maxwell-Boltzmann distribution in the derivation of the chemical rate parameter from molecular collisions (section 4.1.4).

2.5.1 The log-normal distribution

The *log-normal* distribution in a continuous probability distribution of a random variable $\mathcal{Y}$ with a normally distributed logarithm. In other words, if $\mathcal{X} = \ln \mathcal{Y}$ is normally distributed then $\mathcal{Y} = \exp(\mathcal{X})$ has a log-normal distribution. Accordingly $\mathcal{Y}$ can take on only positive real values. Historically, this distribution had several other names the most popular of them being *Galton's distribution* named after the pioneer of statistics in England, Francis Galton or *McAlister's distribution* after the statistician Donald McAlister [235, chap. 14, pp. 207-258].

The log-normal distribution meets the need for modeling empirical data that show frequently observed deviation from the conventional normal distribution: (i) meaningful data are non-negative, (ii) positive skew implying that there are more values above then below the maximum of the probability density function (pdf), and (iii) more obvious meaning of the geometric rather than the arithmetic mean [154, 312]. Despite its obvious usefulness and applicability to problems in science, economics, and sociology the log-normal distribution is not popular among non-statisticians [286].

The log-normal distribution contains two parameters, $\ln \mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}_{>0}$, and is defined on the domain $x \in]0, \infty[$. The density function (pdf) and the cumulative distribution (cdf) are given by (figure 2.13):

$$\begin{aligned} \text{pdf}: \quad f_{\ln \mathcal{N}}(x) &= \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(\ln x - \mu)^2}{2\sigma^2}\Big) \\ \text{cdf}: \quad F_{\ln \mathcal{N}}(x) &= \frac{1}{2}\left(1 + \text{erf}\Big(\frac{\ln x - \mu}{\sqrt{2\sigma^2}}\Big)\right) . \end{aligned} \tag{2.61}$$

By definition the logarithm of the variable $\mathcal{X}$ is normally distributed, and this implies

$$\mathcal{X} \;=\; e^{\mu+\sigma\mathcal{Z}} \;,$$

where $\mathcal{N}$ stand for a standard normal variable. The moments of the log-normal distribution are readily calculated[19]

$$\begin{aligned}
&\text{mean :} && e^{\mu+\sigma^2/2} \;,\\
&\text{median :} && e^{\mu} \;,\\
&\text{mode :} && e^{\mu-\sigma^2} \;,\\
&\text{variance :} && (e^{\sigma^2}-1)\, e^{2\mu+\sigma^2} \;,\\
&\text{skewness :} && (e^{\sigma^2}+2)\,\sqrt{e^{\sigma^2}-1}\,,\ \text{and}\\
&\text{kurtosis :} && e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-6 \;.
\end{aligned} \tag{2.62}$$

The skewness γ_1 is always positive and so is the (excess) kurtosis since $\sigma^2 = 0$ yields $\gamma_2 = 0$, and $\sigma^2 > 0$ implies $\gamma_2 > 0$.

The entropy of the log-normal distribution is

$$H(f_{\ln\mathcal{N}}) \;=\; \frac{1}{2}\Big(1 \,+\, \ln(2\pi\sigma^2) \,+\, 2\,\mu\Big) \;. \tag{2.63}$$

Like the normal distribution has the maximum entropy of all distribution defined on the real axis, $x \in \mathbb{R}$, the log-normal distribution is the maximum entropy probability distribution for a random variable $\mathcal{X}$ for which mean and variance of $\ln\mathcal{X}$ is fixed.

Finally, we mention that the log-normal distribution can be well approximated by a distribution [422]

$$F(x;\,\mu\sigma) \;=\left(\Big(\frac{e^{\mu}}{x}\Big)^{\pi/(\sigma\sqrt{3})} \,+\, 1\right)^{-1}$$

that has integrals that can be expressed in terms of elementary functions.

2.5.2 The χ^2-distribution

The χ^2-distribution also written as chi-squared distribution is one of the most frequently used distribution in inferential statistics for hypothesis testing and construction of confidence intervals. In particular, the χ^2 distributions is applied in the common χ^2-test for the quality of the fit of an empirically deter-

[19] Here and in the following listings for other distributions "kurtosis" stands for excess kurtosis $\gamma_2 = \beta_2 - 3 = \mu_4 \,/\, \sigma^4$.

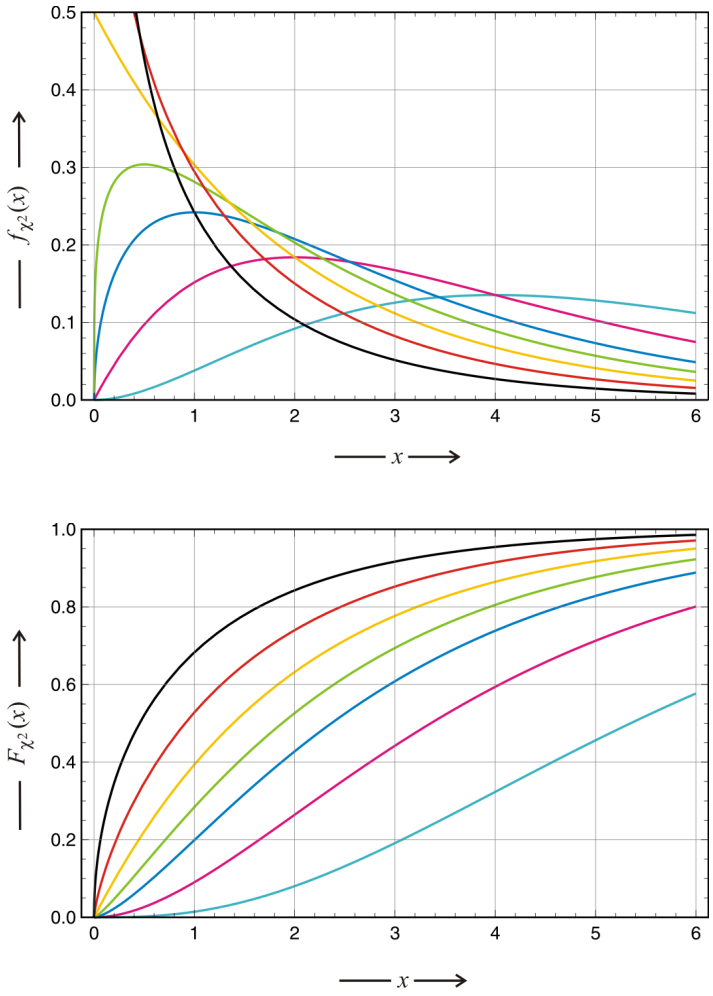


Fig. 2.14 The χ^2 distribution. The chi-squared distribution, χ^2_k , $k \in \mathbb{N}$, is defined on the positive real axis, $x \in [0, \infty[$, with the parameter k called the number of the degrees of freedom, has the probability density (pdf)

$$f_{\chi^2_k}(x) = x^{\frac{k}{2}-1} e^{-\frac{x}{2}} \Big/ \left(2^{\frac{k}{2}} \Gamma\Big(\frac{k}{2}\Big)\right)$$

and the cumulative distribution function (cdf)

$$F_{\chi^2_k}(x) = \gamma\left(\tfrac{k}{2}, \tfrac{x}{2}\right) \Big/ \Gamma\left(\tfrac{k}{2}\right) .$$

Parameter choice and color code: k =1 (black), 1.5 (red), 2 (yellow), 2.5 (green), 3 (blue), 4 (magenta) and 6 (cyan). Although k, the number of degrees of freedom, is commonly restricted to integer values, we show here also the curves for two intermediate values (k=1.5, 2.5).

mined distribution to a theoretical one (section 2.6.2). Many other statistical tests are based on the χ^2-distribution as well.

The chi-squared distribution, χ^2_k,[20] is the distribution of a random variable $\mathcal{Q}$, which is given by the sum of the squares of k independent, standard normal variables with distribution $\mathcal{N}(0,1)$

$$\mathcal{Q} \;=\; \sum_{i=1}^{k} \mathcal{X}_i^2 \;, \tag{2.64}$$

where the only parameter of the distribution, k, is called the number of the *degrees of freedom* being tantamount to the number of independent variables $\mathcal{X}_i$. $\mathcal{Q}$ is defined on the positive real axis (including zero), $x \in [0, \infty[$ and has the following density function and cumulative distribution (figure 2.14):

$$\begin{aligned} \text{pdf}: \quad f_{\chi^2_k}(x) \;&=\; \frac{x^{\frac{k}{2}-1}\, e^{-\frac{x}{2}}}{2^{\frac{k}{2}}} \;, x \in \mathbb{R}_{\geq 0} \quad \text{and} \\ \text{cdf}: \quad F_{\chi^2_k}(x) \;&=\; \frac{\gamma\left(\frac{k}{2}, \frac{x}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \;=\; P\left(\frac{k}{2}, \frac{x}{2}\right) \;. \end{aligned} \tag{2.65}$$

where $\gamma(k,z)$ is the lower incomplete Gamma function and $P(k,z)$ is the regularized Gamma function. The special case with $k = 2$ has the particularly simple form: $F_{\chi^2_2}(x) = 1 - e^{-\frac{x}{2}}$.

The conventional χ^2-distribution is sometimes denoted as *central* χ^2-distribution in order to distinguish it from the *noncentral* χ^2-distribution, which is derived from k independent and normally distributed variables with means μ_i and variances σ_i^2. The random variable

$$\mathcal{Q} \;=\; \sum_{i=1}^{k} \left(\frac{\mathcal{X}_i}{\sigma_i}\right)^2$$

is distributed according to the noncentral χ^2-distribution $\chi^2_k(\lambda)$ with two parameters, k and λ, where $\lambda = \sum_{i=1}^{k} (\mu_i/\sigma_i)^2$ is the *noncentrality* parameter.

The moments of the central χ^2_k-distribution are readily calculated

[20] The chi-squared distribution is sometimes written $\chi^2(k)$ we prefer the subscript since the number of degrees of freedom, the parameter k, specifies the distribution. Often the random variables $\mathcal{X}_i$ fulfil a conservation relation and then the number of independent variables is reduced to $k-1$, and we have χ^2_{k-1} (section 2.6.2).

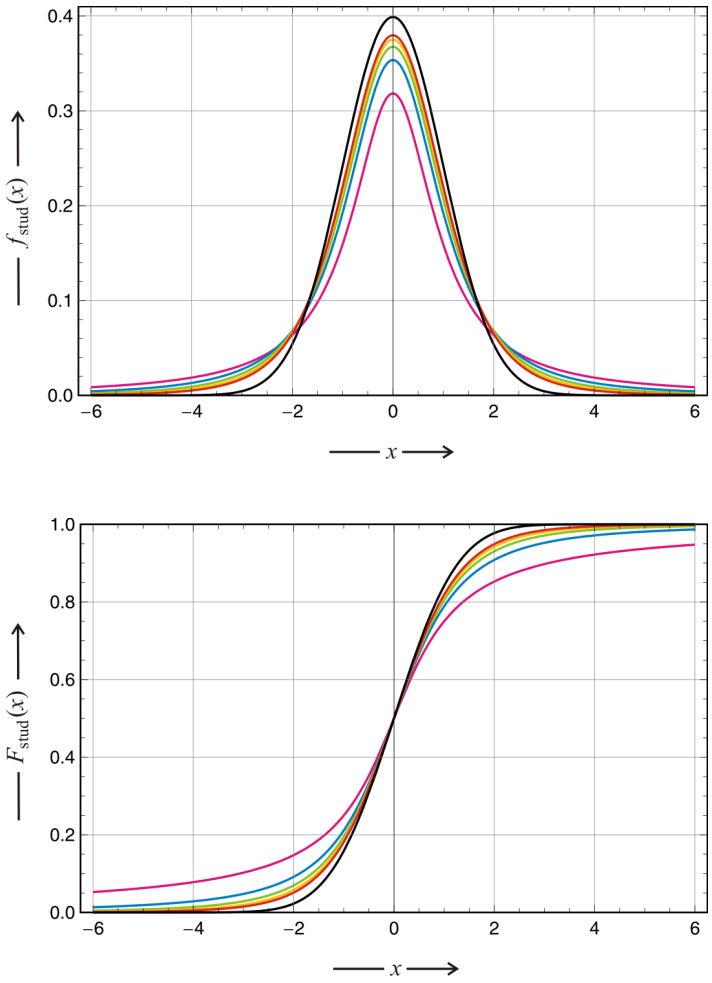


Fig. 2.15 Student's t-distribution. Student's distribution is defined on the real axis, $x \in]-\infty, +\infty[$, with the parameter $r \in \mathbb{N}_{>0}$ called the number of degrees of freedom, has the probability density (pdf)

$$f_{\text{stud}}(x) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi r}\,\Gamma\left(\frac{r}{2}\right)} \left(1 + \frac{x^2}{r}\right)^{\frac{r+1}{2}}$$

and the cumulative distribution function (cdf)

$$F_{\text{stud}}(x) = \tfrac{1}{2} + x\Gamma\left(\frac{r+1}{2}\right) \cdot \frac{{}_2F_1\left(\frac{1}{2}, \frac{r+1}{2}, \frac{3}{2}, -\frac{x^2}{r}\right)}{\sqrt{\pi r}\,\Gamma\left(\frac{r}{2}\right)} \ .$$

The first curve (magenta, $r = 1$) represents the density of the Cauchy-Lorentz distribution (figure 2.18). Parameter choice and color code: r =1 (magenta), 2 (blue), 3 (green), 4 (yellow), 5 (red) and $+\infty$ (black). The black curve representing the limit $r \to \infty$ of Student's distribution is the standard normal distribution.

$$\begin{aligned}
&\text{mean :} && k\ , \\
&\text{median :} && \approx\ k\left(1-\frac{2}{9k}\right)^3\ , \\
&\text{mode :} && \max\{k-2\,,\,0\}\ , \\
&\text{variance :} && 2k\ , \\
&\text{skewness :} && \sqrt{8/k}\,,\ \text{and} \\
&\text{kurtosis :} && 12/k\ .
\end{aligned} \tag{2.66}$$

The skewness γ_1 is always positive and so is the excess kurtosis γ_2. The raw moments $\hat{\mu}_n = \mathrm{E}(\mathcal{Q}^n)$ and the cumulants of the χ^2_k-distribution have particularly simple expressions:

$$\mathrm{E}(\mathcal{Q}^n) \;=\; \hat{\mu}_n \;=\; k(k+2)(k+4)\cdots(k+2n-2) \;=\; 2^n\frac{\Gamma\left(n+\frac{k}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \quad \text{and} \tag{2.67}$$

$$\kappa_n \;=\; 2^{n-1}\,(n-1)!\,k\ . \tag{2.68}$$

The entropy of the χ^2_k-distribution is readily calculated by integration:

$$H(f_{\chi^2}) \;=\; \frac{k}{2} \;+\; \ln\left(2\,\Gamma\big(\frac{k}{2}\big)\right) \;+\; \left(1-\frac{k}{2}\right)\cdot\psi\left(\frac{k}{2}\right)\ , \tag{2.69}$$

where $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\ln\Gamma(x)$ is the digamma function.

The χ^2_k-distribution has a simple characteristic function

$$\phi_{\chi^2}(s) \;=\; (1\;-\;2\mathrm{i}\,s)^{-k/2}\ . \tag{2.70}$$

The moment generating function is defined only for $s < \frac{1}{2}$:

$$M_{\chi^2}(s) \;=\; (1\;-\;2\,s)^{-k/2} \quad \text{for} \quad s\;<\;\frac{1}{2}\ . \tag{2.71}$$

Because of its central importance for tests of significance numerical tables of the χ^2-distribution are found in almost every textbook of mathematical statistics.

2.5.3 Student's t-distribution

Student's t-distribution has a remarkable history. It has been discovered by the famous English statistician William Sealy Gosset who published his works under the pen name *Student* [360]. Gosset was working at the brewery of Arthur Guinness in Dublin, Ireland, where it was forbidden to publish any paper regardless of the contained information, because Guinness was afraid

that trade secrets and other confidential information could be disclosed. Almost all of Gosset's paper including the one describing the t-distribution were published under the pseudonym "Student" [421]. Gosset's work has been known to and was supported by Karl Pearson but it was Ronald Fisher who appreciated the importance of Gosset's work on small samples and made it popular [139].

Student's t-distribution is a family of continuous, normal probability distributions that applies to situations where the sample size is small, the variance is unknown and one wants to derive a reliable estimate of the mean. Student's distribution plays a role in a number of commonly used tests in analyzing statistical data an example being Student's test accessing the significance of differences between two sample means – for example to find out whether or not a difference in mean body height between basketball players and soccer players is significant – or the construction of confidence intervals for the difference between population means. In a way Student's t-distribution is required for *higher order statistics* in the sense of a statistics of statistics, for example, to estimate, how likely it is to find the true mean within a given range around the finite sample mean (section 2.6). In other words, n samples are taken from a population with a normal distribution having fixed but unknown mean and variance, the sample mean and the sample variance are computed from these n points and the t-distribution is the distribution of the location of the true mean relative to the sample mean, calibrated by the sample standard deviation.

To make the meaning of Student's t-distribution precise we assumes n independent random variables $\mathcal{X}_i$, $i = 1, \ldots, n$ drawn from the same population which is normally distributed with mean value $\mathrm{E}(\mathcal{X}_i) = \mu$ and variance $\mathrm{var}(\mathcal{X}_i) = \sigma^2$. Then the sample mean and the unbiased sample variance are the random variables

$$\overline{\mathcal{X}}_n \;= \frac{1}{n}\sum_{i=1}^{n} \mathcal{X}_i \;\text{ and }\; \mathcal{S}_n^2 \;=\; \frac{1}{n-1}\sum_{i=1}^{n} (\mathcal{X}_i - \overline{\mathcal{X}}_n)^2 \;.$$

As follows from Cochran's theorem [70] the random variable $\mathcal{V} = (n-1)\mathcal{S}_n^2/\sigma^2$ follows a χ^2-distribution with $r = n - 1$ degrees of freedom. The deviation of the sample mean from the population mean is properly expressed by the variable

$$\mathcal{Z} \;=\; (\overline{\mathcal{X}}_n - \mu)\,\frac{\sqrt{n}}{\sigma}\;, \tag{2.72}$$

which is the basis for the calculation of z-scores.[21] The variable $\mathcal{Z}$ is normally distributed with mean zero and variance one as follows from the fact that the sample mean $\overline{\mathcal{X}}_n$ obeys a normal distribution with mean μ and variance σ^2/n. In addition, the two random variables $\mathcal{Z}$ and $\mathcal{V}$ are independent, and

[21] In mathematical statistics (section 2.6) the quality of measured data is often characterized by scores. The z-score of a sample corresponds to the random variable $\mathcal{Z}$ (2.72) and it is measured in standard deviations from the population mean as unites.

the *pivotal quantity*[22]

$$\mathcal{T} \doteq \frac{\mathcal{Z}}{\sqrt{\mathcal{V}/(n-1)}} = (\overline{\mathcal{X}}_n - \mu)\,\frac{\sqrt{n}}{\mathcal{S}_n} \tag{2.73}$$

follows a Student's t-distribution, which depends on the degrees of freedom $r = n - 1$ but neither on μ nor on σ.

Student's distribution is a one parameter distribution with r being the number of sample points or the so-called degree of freedom. It is symmetric and bell-shaped like the normal distribution but the tails are heavier in the sense that more values fall further away from the mean. Student's distribution is defined on the real axis, $x \in]-\infty, +\infty[$ and has the following density function and cumulative distribution (figure 2.15):

$$\begin{aligned} \text{pdf}: \quad f_{\text{stud}}(x) &= \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi r}\;\Gamma\left(\frac{r}{2}\right)} \left(1 + \frac{x^2}{r}\right)^{-\frac{r+1}{2}} , \; x \in \mathbb{R} \;\text{ and} \\ \text{cdf}: \quad F_{\text{stud}}(x) &= \frac{1}{2} + x\,\Gamma\left(\frac{r+1}{2}\right) \cdot \frac{{}_2F_1\left(\frac{1}{2}, \frac{r+1}{2}, \frac{3}{2}, -\frac{x^2}{r}\right)}{\sqrt{\pi r}\;\Gamma\left(\frac{r}{2}\right)} . \end{aligned} \tag{2.74}$$

where ${}_2F_1$ is the hypergeometric function. The t-distribution has simple expressions for several special cases:

(i) $r = 1$, Cauchy-distribution: $f(x) = \frac{1}{\pi(1+x^2)}$, $F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan(x)$,

(ii) $r = 2$: $f(x) = \frac{1}{(2+x^2)^{\frac{3}{2}}}$, $F(x) = \frac{1}{2}\left(1 + \frac{x}{\sqrt{2+x^2}}\right)$,

(iii) $r = 3$: $f(x) = \frac{6\sqrt{3}}{\pi(3+x^2)^2}$, $F(x) = \frac{1}{2} + \frac{\sqrt{3}x}{\pi(3+x^2)} + \frac{1}{\pi}\arctan\left(\frac{x}{\sqrt{3}}\right)$,

(iv) $r = \infty$, normal distribution: $f(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}}\,e^{-\frac{x^2}{2}}$, $F(x) = F_{\mathcal{N}}(x)$.

Formally the t-distribution represents an interpolation between the Cauchy-Lorentz distribution (section 2.5.6) and the normal distribution both standardized to mean zero and variance one. In this sense it has a lower maximum and heavier tails than the normal distribution and a higher maximum and less heavy tails than the Cauchy-Lorentz distribution.

The moments of Student's distribution are readily calculated

[22] A pivotal quantity or a pivot is a function of measurable and unmeasurable parameters whose probability distribution does not depend on the unknown parameters.

$$
\begin{aligned}
&\text{mean}: && 0, \text{ for } r > 1, \text{ otherwise undefined}, \\
&\text{median}: && 0\,, \\
&\text{mode}: && 0\,, \\
&\text{variance}: && \begin{cases} \infty & \text{for } \ 1 < r \le 2\,, \\ \frac{r}{r-2} & \text{for } \ r > 2\,, \\ \text{undefined} & \text{otherwise}\,, \end{cases} \\
&\text{skewness}: && 0, \text{ for } r > 3, \text{ otherwise undefined}, \text{ and} \\
&\text{kurtosis}: && \begin{cases} \infty & \text{for } \ 2 < r \le 4\,, \\ \frac{6}{r-4} & \text{for } \ r > 4\,, \\ \text{undefined} & \text{otherwise}\,. \end{cases}
\end{aligned}
\tag{2.75}
$$

The variance of the Student's distribution – provided it is defined – is larger than $\sigma^2 = 1$, the variance of the standard normal distribution. In the limit of infinite degrees of freedom Student's distribution converges to the standard normal distribution and so does the variance: $\sigma^2 = \lim_{r\to\infty} \frac{r}{r-2} = 1$. Student's distribution is symmetric and hance the skewness γ_1 is either zero or undefined, and the (excess) kurtosis γ_2 is undefined or positive and converges to zero in the limit $r \to \infty$.

The raw moments $\hat{\mu}_n = \mathrm{E}(\mathcal{T}^n)$ of the t-distribution have fairly simple expressions:

$$
\mathrm{E}(\mathcal{T}^k) \;=\; \begin{cases} 0 & k \text{ odd}\,,\ 0 < k < r\,, \\ \frac{1}{\sqrt{\pi}\,\Gamma\left(\frac{r}{2}\right)}\, r^{\frac{k}{2}}\, \Gamma\left(\frac{k+1}{2}\right) \Gamma\left(\frac{r-k}{2}\right) & k \text{ even}\,,\ 0 < k < r\,, \\ \text{undefined} & k \text{ odd}\,,\ 0 < r \le k\,, \\ \infty & k \text{ even}\,,\ 0 < r \le k\,. \end{cases}
\tag{2.76}
$$

The entropy of Student's t-distribution is readily calculated by integration:

$$
H(f_{\text{stud}}) \;=\; \frac{k+1}{2}\, \left(\psi\Big(\frac{1+r}{2}\Big) - \psi\Big(\frac{r}{2}\Big) \right) \,+\, \ln\left(\sqrt{r}\, B\Big(\frac{r}{2}, \frac{1}{2}\Big) \right)\,, \tag{2.77}
$$

where $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x} \ln \Gamma(x)$ and $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\,\mathrm{d}t$ are the digamma function and the beta function, respectively.

Student's-distribution has the characteristic function

$$
\phi_{\text{stud}}(s) \;=\; \frac{K_{r/2}(\sqrt{r}\,|s|) \cdot (\sqrt{r}\,|s|)^{r/2}}{2^{\frac{r}{2}-1} \cdot \Gamma\left(\frac{r}{2}\right)} \quad \text{for } \ r > 0\,. \tag{2.78}
$$

where $K_\alpha(x)$ is a modified Bessel function.

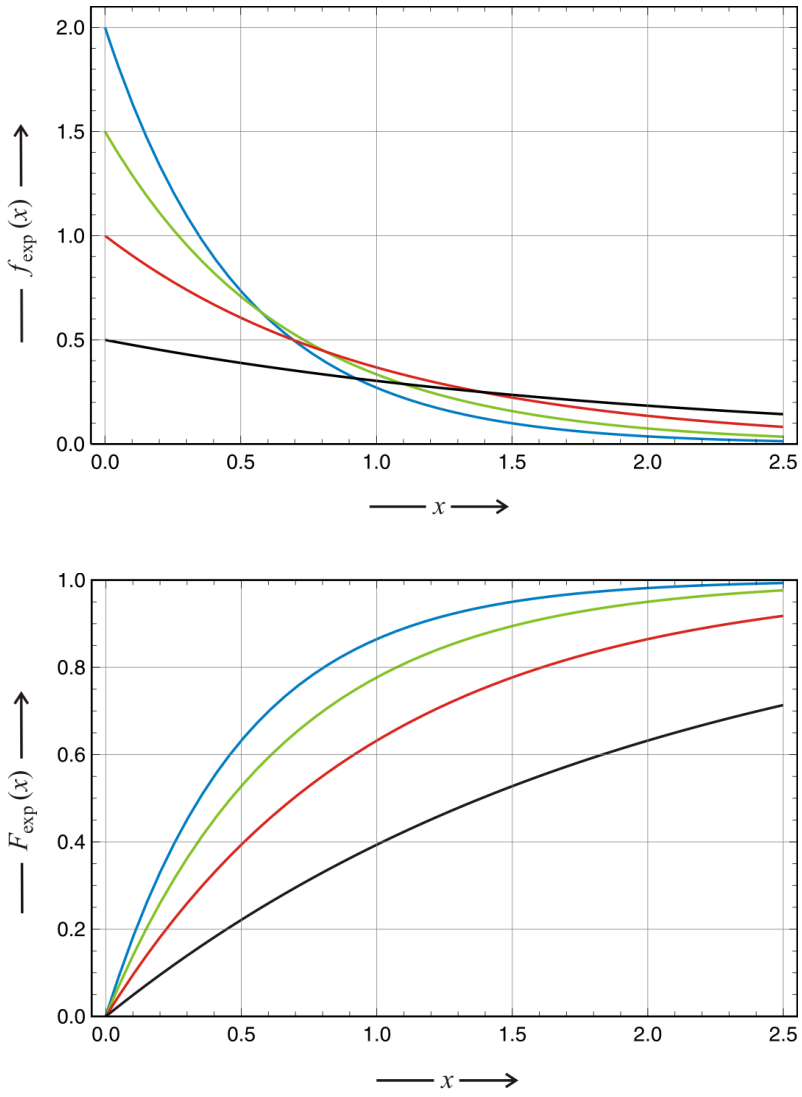


Fig. 2.16 The exponential distribution. The exponential distribution is defined on the real axis including zero, $x \in [0, +\infty[$, with the parameter $\lambda \in \mathbb{R}_{>0}$ called the rate parameter, and has the probability density (pdf)

$$f_{\exp}(x) = \lambda \exp(-\lambda x)$$

and the cumulative distribution function (cdf)

$$F_{\exp}(x) = 1 - \exp(-\lambda x) \ .$$

Parameter choice and color code: λ =0.5 (black), 2 (red), 3 (green), and 4 (blue).

2.5.4 The exponential and the geometric distribution

The *exponential distribution* is a family of continuous probability distributions, which describe the distribution of the time intervals between events in a Poisson process (section 3.2.2.4), which is a process where the number of events in any time interval has a Poisson distribution.[23] The Poisson process is a process where events occur steadily, independently of each other and at a constant average rate $\lambda \in \mathbb{R}_{>0}$, which is the only parameter of the exponential distribution and the Poisson process as well.

The exponential distribution has widespread applications in science and sociology. It describes the time to decay of radioactive atoms, the time to reaction events in irreversible first order processes in chemistry and biology, the waiting times in queues of independently acting customers, the time to failure of components with constant failure rates and other instances.

The exponential distribution is defined on the positive real axis, $x \in [0, \infty[$, with a positive rate parameter $\lambda \in \,]\,0, \infty[$. The density function and cumulative distribution are of the form (figure 2.16):

$$\begin{aligned} \text{pdf}: \quad & f_{\exp}(x) \;=\; \lambda \exp(-\lambda x),\; x \in \mathbb{R}_{>0} \;\text{ and} \\ \text{cdf}: \quad & F_{\exp}(x) \;=\; 1 - \exp(-\lambda x),\; x \in \mathbb{R}_{>0}\;. \end{aligned} \tag{2.79}$$

The moments of exponential distribution are readily calculated

$$\begin{aligned} &\text{mean}: && \lambda^{-1} = \mu\;, \\ &\text{median}: && \lambda^{-1} \ln 2\;, \\ &\text{mode}: && 0\;, \\ &\text{variance}: && \lambda^{-2}\;, \\ &\text{skewness}: && 2\;, \text{ and} \\ &\text{kurtosis}: && 6\;. \end{aligned} \tag{2.80}$$

A commonly used alternative parametrization uses a *survival parameter* $\beta = \lambda^{-1} = \mu$ instead of the rate parameter, and survival is often measured in terms of *half-life*, which is the expectation value of the time when one half of the events have taken place – for example 50 % of the atoms have decayed – and represents just another name for the median: $\overline{\mu} = \beta \ln 2 = \ln 2/\lambda$. The exponential distribution provides an easy to verify test case for the median-mean inequality:

[23] It is important to distinguish the exponential distribution and the class of *exponential families of distributions*, which comprises many other distributions like the normal distribution, the Poisson distribution, the binomial distribution and many others [116].

$$|\mathrm{E}(\mathcal{X}) - \overline{\mu}| \;=\; \frac{1 \,-\, \ln 2}{\lambda} \;<\; \frac{1}{\lambda} \;=\; \sigma \;.$$

The raw moments of the exponential distribution are given simply by

$$\mathrm{E}(\mathcal{X}^n) \;=\; \hat{\mu}_n \;=\; \frac{n!}{\lambda^n} \;. \tag{2.81}$$

Among all probability distribution with the support $[\,0, \infty[$ and mean μ the exponential distribution with $\lambda = 1/\mu$ has the largest entropy (section 2.1.3):

$$H(f_{\exp}) \;=\; 1 \,-\, \log \lambda \;=\; 1 \,+\, \log \mu \;. \tag{2.20'}$$

The moment generation function of the exponential distribution is

$$M_{\exp}(s) \;=\; \left(1 \,-\, \frac{s}{\lambda}\right)^{-1} \;, \tag{2.82}$$

and the characteristic function is

$$\phi_{\exp}(s) \;=\; \left(1 \,-\, \frac{\dot{\imath} s}{\lambda}\right)^{-1} \;. \tag{2.83}$$

Finally, we mention a property of the exponential distribution that makes it unique among all continuous probability distributions: It is *memoryless*. Memorylessness can be encapsulated in an example called "hitchhiker's dilemma": Waiting for hours on a lonely road does not increase the probability of arrival of the next car. Cast into probabilities this means for a random variable $\mathcal{T}$:[24]

$$P(\mathcal{T} > s + t \,|\, T > s) \;=\; P(T > t) \;\; \forall \;\; s, t \,\geq\, 0 \;. \tag{2.84}$$

In other words, the probability of arrival does not change no matter how many events have happened.

The discrete analogue to the exponential distribution is the geometric distribution. Considered is a sequence of independent Bernoulli trials with p being the probability of success and the only parameter of the distribution: $0 < p \leq 1$. The random variable $\mathcal{X} \in \mathbb{N}$ is the number of trials before the first success.

The probability mass function and the cumulative distribution function of the geometric distribution are:

$$\begin{aligned} \mathrm{pmf}: \quad f^{\mathrm{geom}}_{k;p} \;&=\; p \cdot (1-p)^k, \;\; k \in \mathbb{N} \;\; \text{and} \\ \mathrm{cdf}: \quad F^{\mathrm{geom}}_{k;p} \;&=\; 1 \,-\, (1-p)^{k+1}), \;\; k \in \mathbb{N} \;. \end{aligned} \tag{2.85}$$

[24] We remark that memoryless is not tantamount to independence. Independence would require $P(\mathcal{T} > s + t \,|\, T > s) \,=\, P(T > s + t)$.

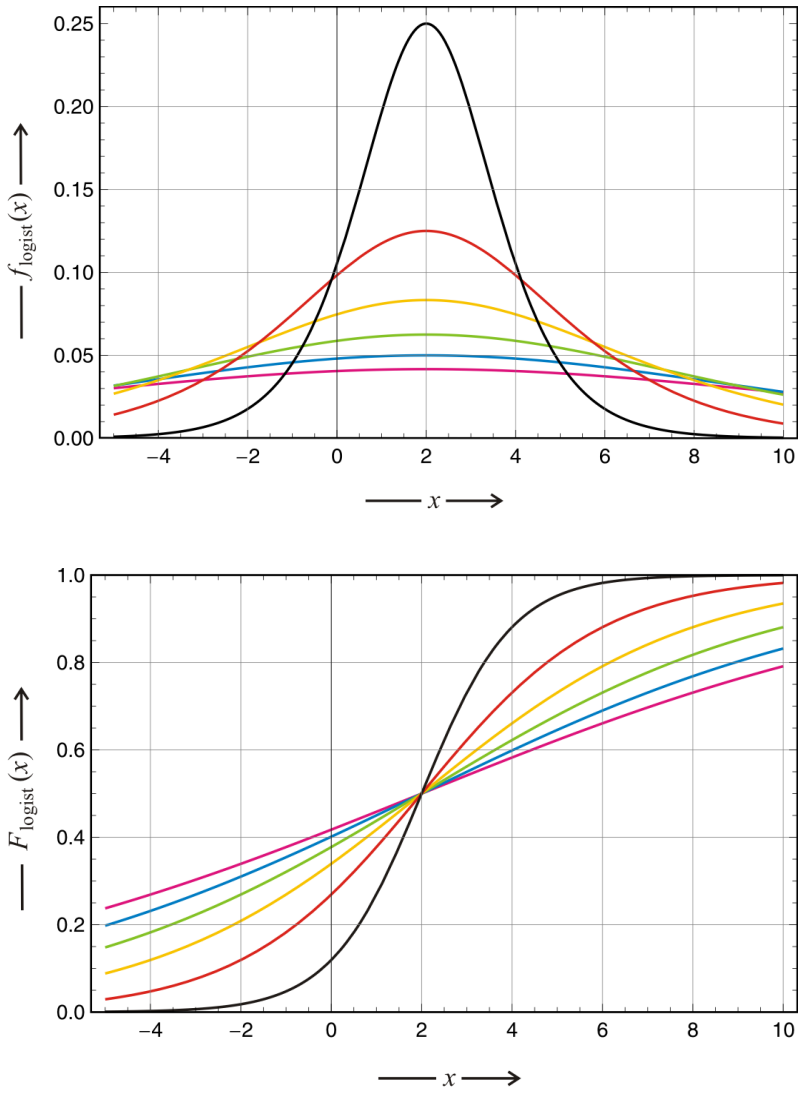


Fig. 2.17 The logistic distribution. The logistic distribution is defined on the real axis, $x \in]-\infty, +\infty[$, with two parameters, the location $\mu \in \mathbb{R}$ and the scale $b \in \mathbb{R}_{>0}$, has the probability density (pdf)

$$f_{\text{logist}}(x) = \frac{e^{-(x-\mu)b}}{b\,(1+e^{-(x-\mu)/b})^2}$$

and the cumulative distribution function (cdf)

$$F_{\text{logist}}(x) = \frac{1}{1+e^{-[x-\mu)/b}} \ .$$

Parameter choice and color code: $\mu = 2$, b =1 (black), 2 (red), 3 (yellow), 4 (green), 5 (blue) and 6 (magenta).

The moments of geometric distribution are readily calculated

$$\begin{aligned}
&\text{mean :} && \frac{1-p}{p} \ , \\
&\text{median :} && \lambda^{-1} \ln 2 \ , \\
&\text{mode :} && 0 \ , \\
&\text{variance :} && \frac{1-p}{p^2} \ , \\
&\text{skewness :} && \frac{2-p}{\sqrt{1-p}} \ , \text{ and} \\
&\text{kurtosis :} && 6 + \frac{p^2}{1-p} \ .
\end{aligned} \tag{2.86}$$

Like the exponential distribution the geometric distribution is lacking memory in the sense of equation (2.84). The information entropy has the form

$$H(f^{\text{geom}}_{k;p}) \ = \ -\frac{1}{p}\Big((1-p)\,\log(1-p) + p\,\log p\Big) \ . \tag{2.87}$$

Finally, we present the moment generating function and the characteristic function of the geometric distribution:

$$M_{\text{geom}}(s) \ = \ \frac{p}{1-(1-p)\exp(s)} \quad \text{and} \tag{2.88}$$

$$\phi_{\text{geom}}(s) \ = \ \frac{p}{1-(1-p)\exp(\dot{\imath}\, s)} \ , \tag{2.89}$$

respectively.

2.5.5 The logistic distribution

The logistic distribution is commonly used as a model for growth with limited resources. It is applied in economics, for example, to model the market penetration of a new product, in biology for population growth in an ecosystem, in agriculture for the expansion of agricultural production or to weight gain in animal fattening. It is a continuous probability distribution with two parameters, the position of the mean μ and the scale b. The cumulative distribution function of the logistic distribution is the *logistic function*.

The logistic distribution is defined on the real axis, $x \in \,]-\infty, \infty[$, with two parameters, the position of the mean $\mu \in \mathbb{R}$ and the scale $b \in \mathbb{R}_{>0}$. The density function and cumulative distribution are of the form (figure 2.17):

$$\begin{aligned} \text{pdf}: \quad f_{\text{logist}}(x) &= \frac{e^{-(x-\mu)b}}{b\,(1+e^{-(x-\mu)/b})^2},\ x \in \mathbb{R}\,,\ \text{and} \\ \text{cdf}: \quad F_{\text{logist}}(x) &= \frac{1}{1+e^{-(x-\mu)/b}},\ x \in \mathbb{R}\ . \end{aligned} \tag{2.90}$$

The moments of logistic distribution are readily calculated

$$\begin{aligned} &\text{mean}: && \mu\ , \\ &\text{median}: && \mu\ , \\ &\text{mode}: && \mu\ , \\ &\text{variance}: && \frac{\pi^2\,b^2}{3}\ , \\ &\text{skewness}: && 0\ ,\ \text{and} \\ &\text{kurtosis}: && \frac{6}{5}\ . \end{aligned} \tag{2.91}$$

A frequently used alternative parametrization uses the variance as parameter, $\sigma = \pi b/\sqrt{3}$ or $b = \sqrt{3}\,\sigma/\pi$. The density and the cumulative distribution can be expressed also in terms of hyperbolic functions

$$f_{\text{logist}}(x) \;=\; \frac{1}{4b}\,\text{sech}^2\left(\frac{x-\mu}{2b}\right) \quad \text{and} \quad F_{\text{logist}}(x) \;=\; \frac{1}{2} + \frac{1}{2}\,\tanh\left(\frac{x-\mu}{2b}\right)\ .$$

The logistic distribution resembles the normal distribution and like Student's distribution the logistic distribution has heavier tails and a lower maximum than the normal distribution. The entropy takes on the simple form

$$H(f_{\text{logist}}) \;=\; \log b \,+\, 2\ . \tag{2.92}$$

The moment generating of the logistic distribution is

$$M_{\text{logist}}(s) \;=\; \exp(\mu s)\,B(1-bs, 1+bs)\ , \tag{2.93}$$

for $|bs| < 1$ and $B(x,y)$ being the Beta function. The characteristic function is

$$\phi_{\text{logist}}(s) \;=\; \frac{\pi bs\exp(\dot{\imath}\,\mu s)}{\sinh(\pi bs)}\ . \tag{2.94}$$

2.5.6 The Cauchy-Lorentz distribution

The Cauchy-Lorentz distribution $\mathcal{C}(\gamma, \vartheta)$ is a continuous distribution with two parameters, the position ϑ and the scale γ. It is named after the French mathematician Augustin Louis Cauchy and the Dutch physicist Hendrik Antoon

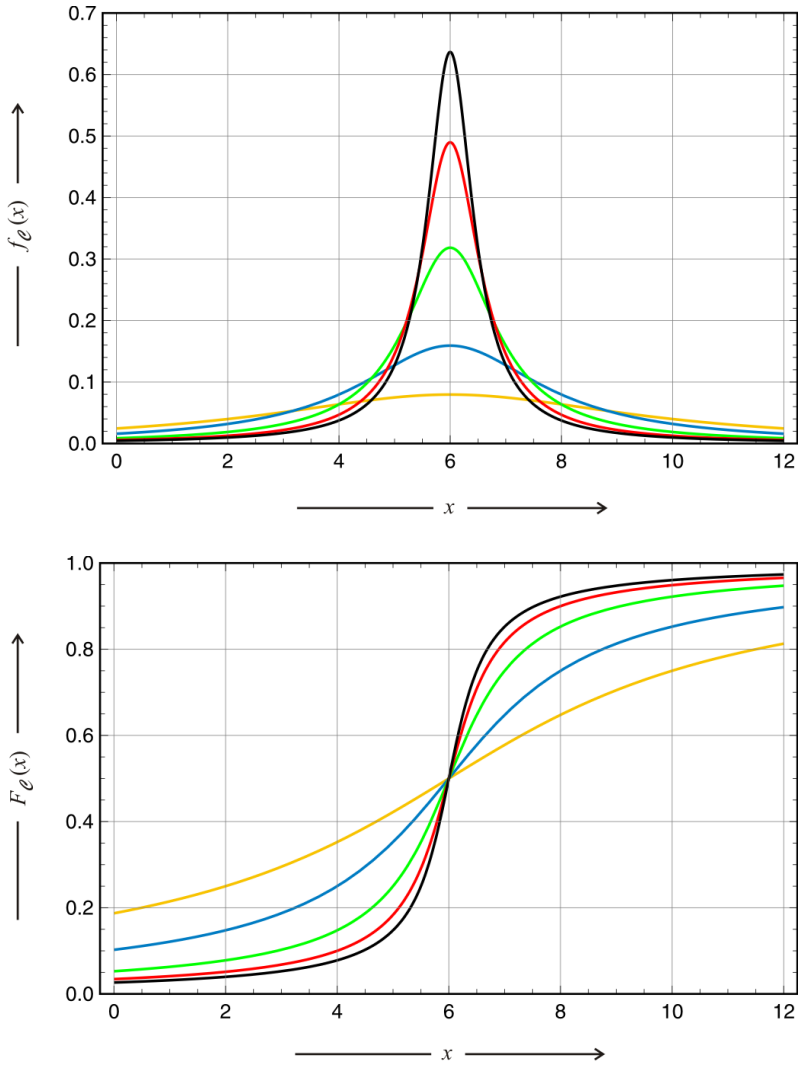


Fig. 2.18 Cauchy-Lorentz density and distribution. In the two plots the Cauchy-Lorentz distribution, $\mathcal{C}(\vartheta, \gamma)$, is shown in from of the probability density

$$f_{\mathcal{C}}(x) = \gamma \;/\; \Big(\pi\big((x-\vartheta)^2 + \gamma^2\big)\Big)$$

and the probability distribution

$$F_{\mathcal{C}}(x) = \tfrac{1}{2} \;+\; \arctan\Big((x-\vartheta)/\gamma\Big) \;/\; \pi\,.$$

Choice of parameters: $\vartheta = 6$ and $\gamma = 0.5$ (black), 0.65 (red), 1 (green), 2 (blue) and 4 (yellow).

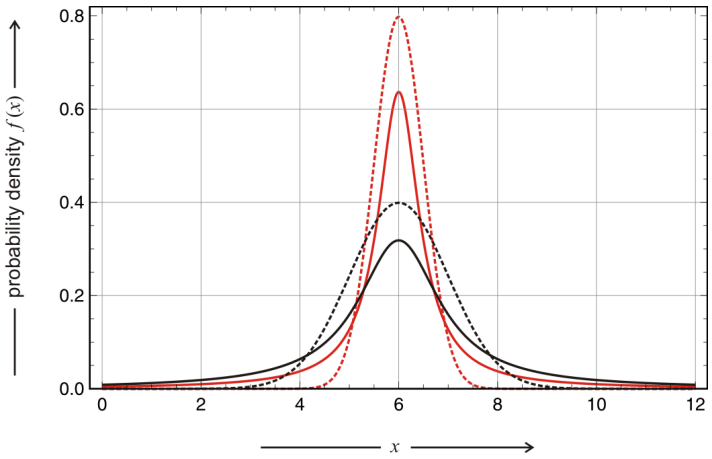


Fig. 2.19 Comparison of Cauchy-Lorentz and normal density. The plots compare the Cauchy-Lorentz density, $\mathcal{C}(\vartheta, \gamma)$ (full lines), and the normal density $\mathcal{N}(\mu, \sigma^2)$ (broken lines). In the flanking regions the normal density decays to zero much faster than the Cauchy-Lorentz density, and this is the cause of the abnormal behavior of the latter. Choice of parameters: $\vartheta = \mu = 6$ and $\gamma = \sigma^2 = 0.5$ (black), 1 (red).

Lorentz and is important in mathematics and in particular in physics where it occurs as the solution to the differential equation for forced resonance. In spectroscopy the Lorentz curve is used for the description of spectral lines that are homogeneously broadened. The Cauchy distribution is a typical heavy-tailed distribution in the sense that larger values of the random variable are more likely to occur in the right tail than in the exponential distribution. Heavy-tailed distributions may also have heavy left tails, or both tails may be heavy as in the Cauchy distribution. As we shall see in section 3.2.5 the Cauchy distribution is a stable distribution and can be partitioned into a sum of Cauchy distributions.

The Cauchy probability density function and the cumulative probability distribution are of the form (figure 2.18)

$$\begin{aligned} \text{pdf}: \quad f_{\mathcal{C}}(x) &= \frac{1}{\pi\,\gamma} \cdot \frac{1}{1 + \left(\frac{x-\vartheta}{\gamma}\right)^2} = \\ &= \frac{1}{\pi} \cdot \frac{\gamma}{(x-\vartheta)^2 + \gamma^2} \quad x \in \mathbb{R} \quad \text{and} \\ \text{cdf}: \quad F_{\mathcal{C}}(x) &= \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x-\vartheta}{\gamma}\right) . \end{aligned} \tag{2.95}$$

The two parameters define the position of the peak, ϑ, and the width of the distribution, γ (figure 2.18). The peak height or amplitude is $1/(\pi\gamma)$. The function $F_{\mathcal{C}}(x)$ can be inverted

$$F_{\mathcal{C}}^{-1}(p) \;=\; \vartheta \,+\, \gamma \,\tan\Big(\pi\big(p - \frac{1}{2}\big)\Big) \tag{2.95'}$$

and we obtain for the quartiles and the median the values: $(\vartheta - \gamma, \vartheta, \vartheta + \gamma)$. As with the normal distribution we define a standard Cauchy distribution $\mathcal{C}(\vartheta, \gamma)$ with $\vartheta = 0$ and $\gamma = 1$, which is identical with Student's t-distribution with one degree of freedom, $r = 1$ (section 2.5.3).

Another remarkable property of the Cauchy distribution concerns the ratio $\mathcal{Z}$ between two independent normally distributed random variables $\mathcal{X}$ and $\mathcal{Y}$ that fulfils a standard Cauchy distribution:

$$\mathcal{Z} \;=\; \frac{\mathcal{X}}{\mathcal{Y}}\,,\; F_{\mathcal{X}} = \mathcal{N}(0,1)\,,\; F_{\mathcal{Y}} = \mathcal{N}(0,1) \quad\Longrightarrow\quad F_{\mathcal{Z}} = \mathcal{C}(0,1)\;,$$

The distribution of the quotient of two random variables is often called the *ratio distribution* and therefore one can say the Cauchy distribution is the *normal ratio distribution*.

Compared to the normal distribution the Cauchy distribution has heavier tails and accordingly a lower maximum (figure 2.19. In this case we cannot use the (excess) kurtosis as an indicator because all moments of the Cauchy distribution are undefined, but we can compute and compare the heights of the standard densities: $f_{\mathcal{C}}(x = \vartheta) = \frac{1}{\pi} \cdot \frac{1}{\gamma}$ and $f_{\mathcal{N}}(x = \mu) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma}$, which yields $f_{\mathcal{C}}(\vartheta) = \frac{1}{\pi}$ and $f_{\mathcal{N}}(\mu) = \frac{1}{\sqrt{2\pi}}$ for $\gamma = \sigma = 1$ with $\frac{1}{\pi} < \frac{1}{\sqrt{2\pi}}$. □The Cauchy distribution has, nevertheless, a defined median and mode, which both coincide with the position of the maximum of the density function, $x = \vartheta$.

The entropy of the Cauchy density is: $H(f_{\mathcal{C}(\vartheta,\gamma)}) = \log\gamma + \log 4$. It cannot be compared with the entropy of the normal distribution in the sense of the maximum entropy principle (section 2.1.3), because this principle refers to distributions with variance σ^2 whereas the variance of the Cauchy distribution is undefined.

The Cauchy distribution has no moment generating function but a characteristic function:

$$\phi_{\mathcal{C}}(s) \;=\; \exp\left(i\,\vartheta\, s \,-\, \gamma\,|s|\right)\,. \tag{2.96}$$

A consequence of the lack of defined moments is that the central limit theorem cannot be applied to a sequence of Cauchy variables. It is can be shown by means of the characteristic function that the mean of a sequence of independent and identically distributed random variables with standard Cauchy distribution, $\mathcal{S} = \sum_{i=1}^{n} \mathcal{X}_i/n$ has the sam standard Cauchy distribution and is not normally distributed as the central limit theorem predicts.

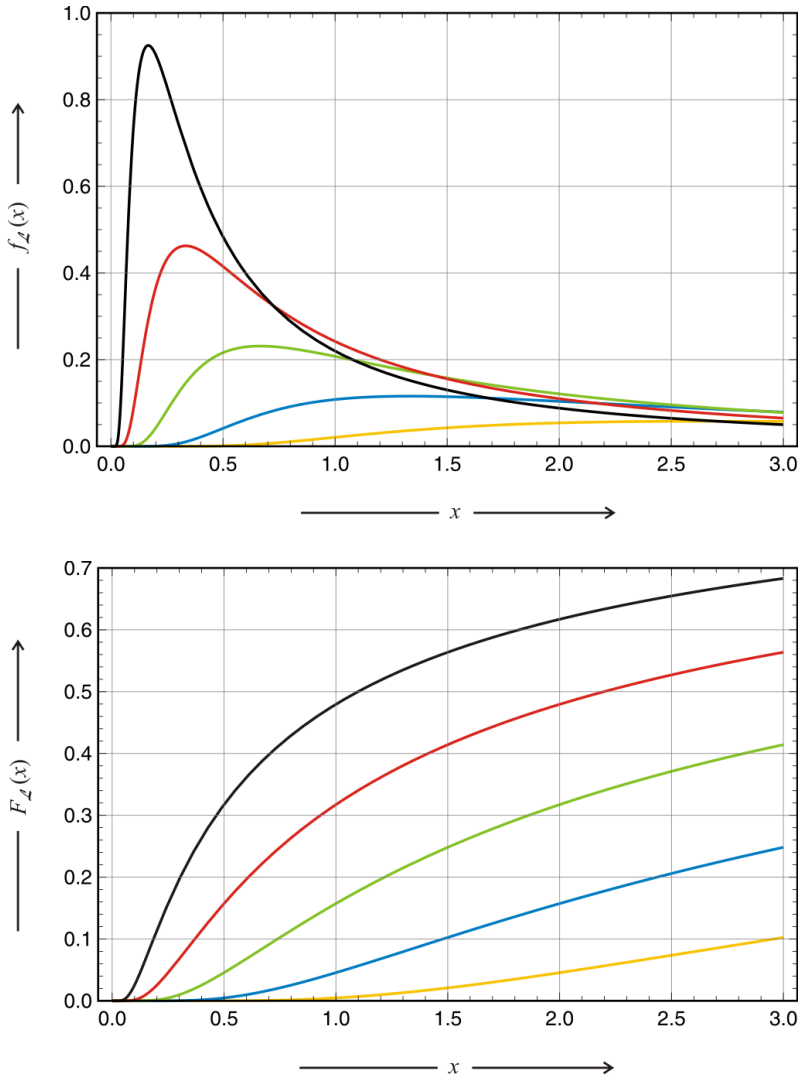


Fig. 2.20 Lévy density and distribution. In the two plots the Lévy distribution, $\mathcal{L}(\vartheta, \gamma)$, is shown in from of the probability density

$$f_{\mathcal{L}}(x) = \sqrt{\frac{\gamma}{2\pi}} \exp\left(-\frac{\gamma}{2(x-\vartheta)}\right) \,/\, (x-\vartheta)^{3/2}$$

and the probability distribution

$$F_{\mathcal{L}}(x) = \operatorname{erfc}\left(\sqrt{\frac{\gamma}{2(x-\vartheta)}}\right).$$

Choice of parameters: $\vartheta = 0$ and $c = 0.5$ (black), 1 (red), 2 (green), 4 (blue) and 8 (yellow).

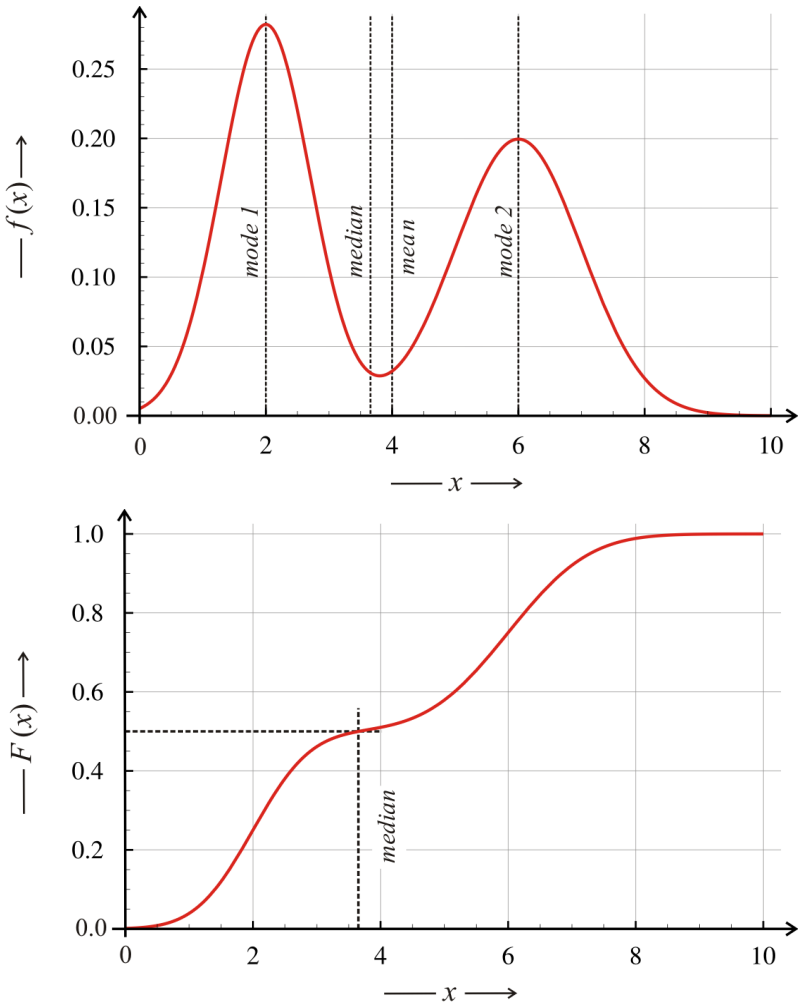


Fig. 2.21 A bimodal probability density. The figure illustrates a bimodal distribution modeled as a superposition of two normal distributions (2.99) with $\alpha = 1/2$ and different values for mean and variance ($\nu_1 = 2, \sigma_1^2 = 1/2$) and ($\nu_2 = 6, \sigma_2^2 = 1$): $f(x) = (\sqrt{2}e^{-(x-2)^2} + e^{-(x-6)^2/2}) \;/\; (2\sqrt{2\pi})$. The upper part shows the probability density corresponding to the two modes $\widetilde{\mu}_1 = \nu_1 = 2$ and $\widetilde{\mu}_2 = \nu_2 = 6$. Median $\bar{\mu} = 3.65685$ and mean $\mu = 4$ are situated near the density minimum between the two maxima. The lower part presents the cumulative probability distribution, $F(x) = \frac{1}{4}\left(2 + \mathrm{erf}(x-2) + \mathrm{erf}\left(\frac{x-6}{\sqrt{2}}\right)\right)$, as well as the construction of the median. The variances in this example are: $\hat{\mu}_2 = 20.75$ and $\mu_2 = 4.75$.

2.5.7 The Lévy distribution

The Lévy distribution $\mathcal{L}(\gamma, \vartheta)$ is a continuous one-sided probability distribution, which is defined for values of the variable x that are larger or equal a shift parameter ϑ: $x \in [\vartheta, \infty[$. It is a special case of the *inverse gamma distribution* and belongs together with the normal and the Cauchy distribution the class of analytically accessible *stable distributions*.

The Lévy probability density function and the cumulative probability distribution are of the form (figure 2.20)

$$\begin{aligned} \text{pdf}: \quad f_{\mathcal{L}}(x) &= \sqrt{\frac{\gamma}{2\pi}} \cdot \frac{1}{(x-\vartheta)^{3/2}} \exp\left(-\frac{\gamma}{2(x-\vartheta)}\right), \; x \in [\vartheta, \infty[, \\ \text{cdf}: \quad F_{\mathcal{L}}(x) &= \operatorname{erfc}\left(\sqrt{\frac{\gamma}{2(x-\vartheta)}}\right). \end{aligned} \tag{2.97}$$

The two parameters $\vartheta \in \mathbb{R}$ and $\gamma \in \mathbb{R}_{>0}$ are the location of $f_{\mathcal{L}}(x) = 0$ and the scale parameter. Mean and variance of the Lévy distribution are infinite, skewness and kurtosis undetermined. For $\vartheta = 0$ the mode of the distribution appears at $\widetilde{\mu} = \gamma/3$ and the median takes on the value $\bar{\mu} = \gamma/\big(2(\operatorname{erfc}^{-1}(1/2))^2\big)$.

The entropy of the Lévy distribution is

$$H\big(f_{\mathcal{L}}(x)\big) = \frac{1 + 3\gamma + \ln(16\pi\gamma^2)}{2} \quad \text{with } \gamma \text{ being Euler's constant,}$$

and the characteristic function

$$\phi_{\mathcal{L}}(s) = \exp\left(\dot{\imath}\,\vartheta\, s - \sqrt{-2\dot{\imath}\,\gamma\, s}\right), \tag{2.98}$$

is the only defined generating function. We shall encounter the Lévy distribution when Lévy processes are discussed in section 3.2.5.

2.5.8 Bimodal distributions

As the name of the bimodal distribution indicates that the density function $f(x)$ has two maxima. It arises commonly as a mixture of two unimodal distribution in the sense that the bimodally distributed random variable $\mathcal{X}$ is defined as

$$P(\mathcal{X}) \;=\; \begin{cases} P(\mathcal{X} \,=\, \mathcal{Y}_1) \;=\; \alpha \;\text{ and} \\ P(\mathcal{X} \,=\, \mathcal{Y}_2) \;=\; (1-\alpha)\,. \end{cases}$$

Bimodal distributions commonly arise from statistics of populations that are split into two subpopulations with sufficiently different properties. The sizes of weaver ants give rise to a bimodal distributions because of the existence of two classes of workers [457]. In case the differences are too small as in case of the combined distribution of body heights for men and women monomodality is observed [390].

As an illustrative model we choose the superposition of two normal distributions with different means and variances (figure 2.21). The probability density for $\alpha = 1/2$ is then of the form:

$$f(x) \;=\; \frac{1}{2\sqrt{2\pi}}\Big(e^{\frac{-(x-\nu_1)^2}{2\sigma_1^2}}\Big/\sqrt{\sigma_1^2} \,+\, e^{\frac{-(x-\nu_2)^2}{2\sigma_2^2}}\Big/\sqrt{\sigma_2^2}\Big)\,. \tag{2.99}$$

The cumulative distribution function is readily obtained by integration. As in the case of the normal distribution the result is not analytical but formulated in terms of the error function, which is available only numerically through integration:

$$F(x) \;=\; \frac{1}{4}\left(2 \,+\, \mathrm{erf}\left(\frac{x-\nu_1}{\sqrt{2\sigma_1^2}}\right) \,+\, \mathrm{erf}\left(\frac{x-\nu_2}{\sqrt{2\sigma_2^2}}\right)\right)\,. \tag{2.100}$$

In the numerical example shown in figure 2.21 the distribution function shows two distinct steps corresponding to the maxima of the density $f(x)$.

As an exercise first an second moments of the bimodal distribution can be readily computed analytically. The results are:

$$\hat{\mu}_1 = \mu = \frac{1}{2}\,(\nu_1+\nu_2)\,,\;\; \mu_1 = 0 \;\text{ and}$$

$$\hat{\mu}_2 = \frac{1}{2}\,(\nu_1^2+\nu_2^2) + \frac{1}{2}\,(\sigma_1^2+\sigma_2^2)\,,\;\; \mu_2 = \frac{1}{4}\,(\nu_1-\nu_2)^2 + \frac{1}{2}\,(\sigma_1^2+\sigma_2^2)\,.$$

The centered second moment illustrates the contributions to the variance of the bimodal density. It is composed of the sum of the variances of the subpopulations and the square of the difference between the two means, $(\nu_1 - \nu_2)^2$.

2.6 Mathematical statistics

Mathematical statistics provides the bridge between probability theory and the analysis of real data, which inevitably represent incomplete since finite samples. Nevertheless, it turned out very appropriate to use infinite samples as a reference (section 1.3). Large sample theory and in particular the law of large numbers (section 2.4.2) deal with the asymptotic behavior of series of samples with increasing size. Although mathematical statistics is a discipline in its own right and would require a separate monograph for a comprehensive presentation, a brief account one three basic concepts, which are of general importance for every scientist will be included here.[25]

First we shall be concerned with approximations to moments derived from finite samples. In practice, we can collect data for all points of the sample space Ω only in very few exceptional cases. Otherwise exhaustive measurements are impossible and we have to rely on limited samples as they are obtained in physics through experiments or in sociology through opinion polls. As an example for the evaluation of the justification of assumptions we introduce Pearson's chi-squared test and finally we illustrate statistical inference by means of an example applying Bayes' theorem.

2.6.1 Sample moments

As we did before for complete sample spaces, we evaluate functions Z from incomplete random samples $(\mathcal{X}_1, \ldots, \mathcal{X}_n)$ and obtain as output random variables $\mathcal{Z} = Z(\mathcal{X}_1, \ldots, \mathcal{X}_n)$. Similarly we compute sample expectation values, also called sample means, sample variances, sample standard deviations and other quantities as estimators from limited sets of data, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. They are calculated in the same way as if the sample set would cover the entire sample space. In particular we compute the *sample mean*

$$m \;=\!=\; \widetilde{\mu} \;=\; \frac{1}{n} \sum_{i=1}^{n} x_i \tag{2.101}$$

and the moments around the sample mean. For the *sample variance* we obtain

[25] For the reader who is interested in more details on mathematical statistics we recommend the classical textbook by the Polish mathematician Marek Fisz [145] and the comprehensive treatise by Stuart and Ord [419, 420], which is a new edition of Kendall's classic on statistics. A text that is useful as an not too elaborate introduction is found in [212], the monograph [73] is particularly addressed to experimentalists practicing statistics, and a great variety of other and equally well suitable texts are, of course, available in the rich literature on mathematical statistics.

$$m_2 \; = = \; \frac{1}{n}\sum_{i=1}^{n} x_i^2 \; - \; \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 \; , \tag{2.102}$$

and for the third and fourth moments after some calculations

$$m_3 \; = \frac{1}{n}\sum_{i=1}^{n} x_i^3 \; - \; \frac{3}{n^2}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{n} x_j^2\right) \; + \; \frac{2}{n^3}\left(\sum_{i=1}^{n} x_i\right)^3 \tag{2.103a}$$

$$\begin{aligned} m_4 \; &= \frac{1}{n}\sum_{i=1}^{n} x_i^4 \; - \; \frac{4}{n^2}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{n} x_j^3\right) \; + \\ &+ \; \frac{6}{n^3}\left(\sum_{i=1}^{n} x_i\right)^2\left(\sum_{j=1}^{n} x_j^2\right) \; - \; \frac{3}{n^4}\left(\sum_{i=1}^{n} x_i\right)^4 \; . \end{aligned} \tag{2.103b}$$

These naïve estimators, m_i $(i = 2, 3, 4, \ldots)$, contain a bias because the exact expectation value μ around which the moments are centered is not known and has to be approximated by the sample mean m. For the variance we illustrate the systematic deviation by calculating a correction factor known as Bessel's correction. This correction, however, is more properly attributed to Carl Friedrich Gauss [244, part 2, p.161]. In order to obtain an expectation value for the sample moments we repeat drawing of samples with n elements and denote their expectation values by $< m_i >$.[26] In particular we have

$$\begin{aligned} m_2 \; &= \; \frac{1}{n}\sum_{i=1}^{n} x_i^2 \; - \; \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 \; = \\ &= \; \frac{1}{n}\sum_{i=1}^{n} x_i^2 \; - \; \frac{1}{n^2}\left(\sum_{i=1}^{n} x_i^2 \; + \sum_{i,j=1,\, i\neq j}^{n} x_i\, x_j\right) \; = \\ &= \; \frac{n-1}{n^2}\sum_{i=1}^{n} x_i^2 \; - \; \frac{1}{n^2}\sum_{i,j=1,\, i\neq j}^{n} x_i\, x_j \; . \end{aligned}$$

The expectation value is now of the form

[26] It is important to note that $< m_i >$ is the expectation value of an average over a finite sample, whereas the genuine expectation value refers to the entire sample space. In particular, we find

$$< m >= \left\langle \tfrac{1}{n}\textstyle\sum_{i=1}^{n} x_i \right\rangle = \mu = \alpha_1 \; ,$$

where μ is the first (raw) moment. For the higher moments the situation is more complicated and requires some care (see text).

$$< m_2 > = \frac{n-1}{n} \left\langle \frac{1}{n} \sum_{i=1}^{n} x_i^2 \right\rangle - \frac{1}{n^2} \left\langle \sum_{i,j=1,\, i \neq j}^{n} x_i \, x_j \right\rangle ,$$

and by using $< x_i x_j >=< x_i >< x_j >=< x_i >^2$ we find

$$\begin{aligned} < m_2 > &= \frac{n-1}{n} \left\langle \frac{1}{n} \sum_{i=1}^{n} x_i^2 \right\rangle - \frac{n(n-1)}{n^2} \left\langle \sum_{i=1}^{n} x_i \right\rangle^2 = \\ &= \frac{n-1}{n} \, \alpha_2 - \frac{n(n-1)}{n^2} \, \mu^2 = \frac{n-1}{n} \, (\alpha_2 - \mu^2) , \end{aligned}$$

where $\alpha_2 = \hat{\mu}_2$ is the second raw moment or second moment about zero. Using the identity $\alpha_2 = \mu_2 + \mu^2$ we find for the unbiased sample variance

$$< m_2 > = \frac{n-1}{n} \, \mu_2 \quad \text{and} \quad \widetilde{\text{var}}(\mathbf{x}) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m)^2 . \tag{2.104}$$

Further useful measures of correlation between pairs of random variables can be derived straightforwardly: (i) the unbiased *sample covariance*

$$\mathcal{M}_{\mathcal{XY}} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m) \, (y_i - m) , \tag{2.105}$$

and (ii) the *sample correlation coefficient*

$$\mathcal{R}_{\mathcal{XY}} = \frac{\sum_{i=1}^{n} (x_i - m) \, (y_i - m)}{\sqrt{(\sum_{i=1}^{n} (x_i - m)^2) \, (\sum_{i=1}^{n} (y_i - m)^2)}} . \tag{2.106}$$

For practical purposes Bessel's correction is often unimportant when the data sets are sufficiently large but the recognition of the principle is important in particular for statistical properties more involved than variances. Sometimes a problem is encountered in cases where the second moment of a distribution, μ_2, does not exist, which means it diverges. Then, computing variances from incomplete data sets is also unstable and one may choose the *mean absolute deviation*,

$$\mathcal{D}(\mathcal{X}) = \frac{1}{n} \sum_{i=1}^{n} |\mathcal{X}_i - m| , \tag{2.107}$$

as a measure for the width of the distribution [374, pp.455-459], because it is commonly more robust than variance or standard deviation.

Ronald Fisher conceived k-statistics in order to derive estimators for the moments of finite samples [140]. The cumulants of a probability distribution are derived as expectation values, $< k_i >= \kappa_i$, of finite set cumulants calculated in the same way as the complete sample set analogues [245, pp.99-100]. The first four terms of k-statistics for n sample points are

$$
\begin{aligned}
k_1 &= m \ , \\
k_2 &= \frac{n}{n-1}\, m_2 \ , \\
k_3 &= \frac{n^2}{(n-1)(n-2)}\, m_3 \quad \text{and} \\
k_4 &= \frac{n^2 \left((n+1)\, m_4 \ - \ 3\,(n-1)\, m_2{}^2 \right)}{(n-1)(n-2)(n-3)} \ ,
\end{aligned}
\tag{2.108}
$$

which can be derived by inversion of the well known relationships

$$
\begin{aligned}
< m > &= \mu \ , \\
< m_2 > &= \frac{n-1}{n}\, \mu_2 \ , \\
< m_3 > &= \frac{(n-1)(n-2)}{n^2}\, \mu_3 \ , \\
< m_2^2 > &= \frac{(n-1)\Big((n-1)\, \mu_4 \ + \ (n^2 - 2n + 3)\, \mu_2{}^2 \Big)}{n^3} \ , \quad \text{and} \\
< m_4 > &= \frac{(n-1)\Big((n^2 - 3n + 3)\, \mu_4 \ + \ 3\,(2n-3)\, \mu_2{}^2 \Big)}{n^3} \ .
\end{aligned}
\tag{2.109}
$$

The usefulness of these relations becomes evident in various applications.

The statistician computes moments and other functions from his empirical data sets, which is almost always non-exhaustive, for example $\{x_1, \ldots, x_n\}$ or $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ by means of the equations (2.101) and (2.104) to (2.106). The underlying assumption, of course, is that the values of the empirical functions converge to the corresponding exact moments as the random sample increases and the theoretical basis for this assumption is provided by the law of large numbers.

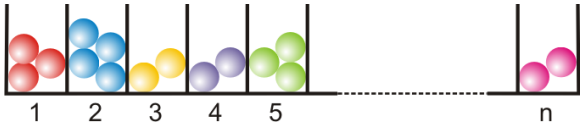


Fig. 2.22 Definition of cells for the application of the χ^2-square test. The space of possible outcomes of recordings is partitioned into n cells, which correspond to features of classification. As an example one could group animals into males and females or scores according to the numbers on the top face of a rolled die. The characteristics of classification are visualized by different colors.

2.6.2 Pearson's chi-squared test

The main issue of mathematical statistics, however, is not so much to compute approximations to the moments but – as it has always been and still – the development of independent tests that allow for the derivation of information on the appropriateness of models and the quality of data. Predictions on the reliability of the computed values are made by means of a great variety of tools. We dispense from details, which are extensively treated in the literature [146, 419, 420]. Karl Pearson conceived a test in 1900 [364], which became popular under the name chi-squared test. This test has also been used by Ronald Fisher when he analyzed Gregor Mendel's data on genetics of the garden pea *Pisum sativum* and we shall use the data given in table 1.1 to illustrate the application of the chi-squared test.

The formula of Pearson's test is made plausible by means of a simple example [213, pp. 407-414]: A random variable $\mathcal{Y}_1$ is binomially distributed according to $B_k(n, p_1)$ with expectation value $\mathrm{E}(\mathcal{Y}_1) = n\,p_1$ and variance $\sigma_1^2 = n\,p_1(1-p_1)$ (section 2.3.2) and then, following the central limit theorem the random variable

$$\mathcal{Z} \;=\; \frac{\mathcal{Y}_1 \;-\; np_1}{\sqrt{np_1(1-p_1)}}$$

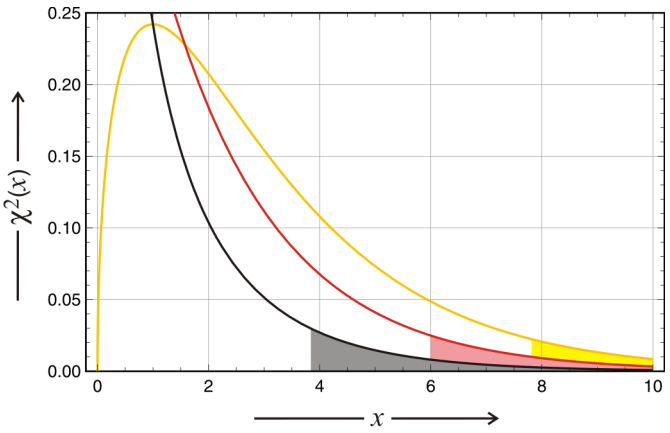


Fig. 2.23 The p-value in significance test of null hypothesis. The figure illustrates the definition of the p-value. The three curves represent the χ_k^2 probability densities with the parameters $k = 1$ (black), 2 (red), and 3 (yellow). The three specific $x_k(\alpha)$-values are shown for the critical p-value with $\alpha = 0.05$ are shown: for $k = 1$ we find $x_1(0.05) = 3.84146$, for $k = 2$ we obtain $x_2(0.05) = 5.99146$, and for $k = 3$ eventually $x_3(0.05)$. The hatched areas show the range of values of the random variable $\mathcal{Q}$ that are more extreme than the predefined critical p-value, which is defined as the cumulative probability within the indicated areas that were defined by $\alpha = 0.05$. If the p-value for an observed data set fulfils $p < \alpha$, the null hypothesis is rejected.

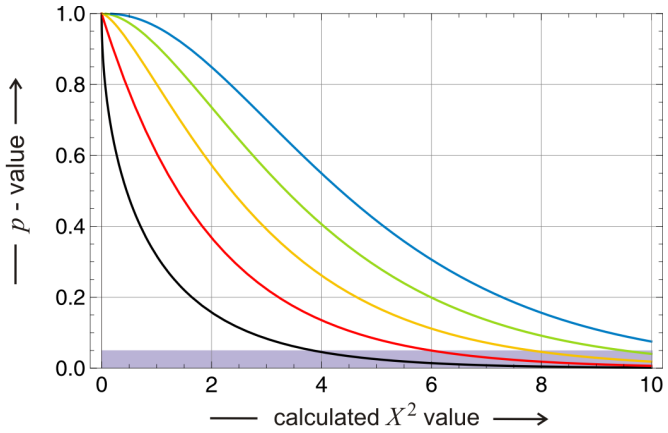


Fig. 2.24 Calculation of the p-value in significance test of null hypothesis. The figure shows the p-values from equation (2.115) as a function of the calculated values of X_k^2 for the k-values 1 (black), 2 (red), 3 (yellow), 4 (green), and 5 (blue). The highlighted area at the bottom of the figure shows the range where the null hypothesis is rejected.

has a standardized binomial distribution, which approximates $\mathcal{N}(0,1)$ for sufficiently large n (section 2.4.1). A second random variable is $\mathcal{Y}_2 = n - \mathcal{Y}_1$ with expectation value $\mathrm{E}(\mathcal{Y}_2) = n\,p_2$ and variance $\sigma_2^2 = \sigma_1^2 = n\,p_2(1-p_2) = n\,p_1(1-p_1)$, since $p_2 = (1-p_1)$. The sum $\mathcal{Z}^2 = \mathcal{Y}_1^2 + \mathcal{Y}_2^2$ is approximately χ^2 distributed:

$$\mathcal{Z}^2 \;=\; \frac{(\mathcal{Y}_1 - np_1)^2}{np_1(1-p_1)} \;=\; \frac{(\mathcal{Y}_1 - np_1)^2}{np_1} + \frac{(\mathcal{Y}_2 - np_2)^2}{np_2} \quad \text{since}$$

$$(\mathcal{Y}_1 - np_1)^2 \;=\; \Big(n \,-\, \mathcal{Y}_1 \,-\, n(1-p_1)\Big)^2 \;=\; (\mathcal{Y}_2 - np_2)^2 \;.$$

We can now rewrite the expression by introducing the expectation values

$$\mathcal{Q}_1 \;=\; \sum_{i=1}^{2} \frac{\Big(\mathcal{Y}_i - \mathrm{E}(\mathcal{Y}_i)\Big)^2}{\mathrm{E}(\mathcal{Y}_i)} \;,$$

and indicating the number of independent random variables as a subscript. Provided all products $n\,p_i$ are sufficiently large – a conservative estimate would be $np_i \geq 5\,\forall\, i$ – the quantity $\mathcal{Q}_1$ has an approximate chi-squared distribution with one degree of freedom: χ_1^2.

The generalization to an experiment with k mutually exclusive and exhaustive outcomes $A_1, A_2, \ldots, A_k$ of the variables $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_k$, is straightforward. All variables $\mathcal{X}_i$ are assumed to have finite mean μ_i and finite variance σ_i^2 such that central limit theorem applies and the distribution for large n

converges to the normal distribution $\mathcal{N}(0,1)$. We define the probability to obtain the result A_i by $P(A_i) = p_i$, by conservation of probabilities we have $\sum_{i=1}^{k} p_i = 1$, and thus one variable is lacking independence and we choose it to be $\mathcal{X}_k$:

$$\mathcal{X}_k \;=\; n \,-\, \sum_{i=1}^{k-1} \mathcal{X}_i \;. \tag{2.110}$$

The joint distribution of $k-1$ variables $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_{k-1}$ has then the joint probability mass function (pmf)

$$f(x_1, x_2, \ldots, x_{k-1}) \;=\; P(\mathcal{X}_1 = x_1, \mathcal{X}_2 = x_2, \ldots, \mathcal{X}_{k-1} = x_{k-1}) \;.$$

Next we consider n independent trials yielding x_1 times A_1, x_2 times $A_2, \ldots,$ and x_k times A_k, where a particular outcome has the probability

$$P(\mathcal{X}_1 = x_1, \mathcal{X}_1 = x_1, \ldots, \mathcal{X}_{k-1} = x_{k-1}) \;=\; p_1^{x_1} \cdot p_2^{x_2} \cdot \ldots \cdot p_k^{x_k} \quad \text{with}$$

the frequency factor or statistical weight

$$g(x_1, x_2, \ldots, x_k) \;=\; \binom{n}{x_1, x_2, \ldots, x_k} \;=\; \frac{n!}{x_1!, x_2!, \ldots, x_k!} \;,$$

and eventually we find for the pmf

$$\begin{aligned} f(x_1, x_2, \ldots, x_{k-1}) \;&=\; g(x_1, x_2, \ldots, x_k) \cdot P(\mathcal{X}_1 = x_1, \mathcal{X}_2 = x_2, \ldots, \mathcal{X}_{k-1} = x_{k-1}) \;= \\ &=\; \frac{n!}{x_1!, x_2!, \ldots, x_k!}\, p_1^{x_1} \cdot p_2^{x_2} \cdot \ldots \cdot p_k^{x_k} \;, \end{aligned} \tag{2.111}$$

with the two restrictions $x_k = n - \sum_{i=1}^{k-1} x_i$ and $p_k = 1 - \sum_{i=1}^{k-1} p_i$. Pearson's construction follows the lines we have shown for the binomial with $k = 2$ and yields under consideration of equation 2.110:

$$\mathcal{Q}_{k-1}(n) \;=\; X^2_{k-1}(n) \;=\sum_{i=1}^{k} \frac{\Big(\mathcal{X}_i \,-\, \mathrm{E}(\mathcal{X}_i)\Big)^2}{\mathrm{E}(\mathcal{X}_i)} \;. \tag{2.112}$$

The sum of squares $\mathcal{Q}_{k-1}(n)$ in (2.112) is called Pearson's *cumulative test statistic*. It has an approximate chi-squared distribution with k-1 degrees of freedom, χ^2_{k-1},[27] and again if n is sufficiently large to fulfil $n \cdot p_1 \geq 5 \;\forall\, i$ the distributions are close enough for most practical purposes.

In order to be able to test hypotheses we divide our sample space into k cells and record observations falling into individual cells. In essence, these cells $\mathcal{C}_i$ are tantamount to the outcomes A_i but we can define them to be completely general, for example collecting all instances that falls in a certain

[27] We indicate the expected converge in the sense of the central limit theorem by choosing the symbol X^2_{k-1} for the finite n expression with $\lim_{n\to\infty} X^2_{k-1}(n) = \chi^2_{k-1}$.

range. At the end of the registration period the number of observations is n and partitioning into the instances that were recorded in the cell $\mathcal{C}_i$ is ν_i with $\sum_{i=1}^{k} \nu_i = n$. Equation (2.112) is now applied to test a (null) hypothesis H_0 against empirically registered values for the different outcomes,

$$H_0 : \quad E_i^{(0)}(\mathcal{X}_i) \, = \, \varepsilon_{i0} \, , \; i = 1, \ldots, k \; . \tag{2.113}$$

In other words, the null hypothesis predicts the distribution of score values falling into the cells $\mathcal{C}_i$ to be ε_{i0} $(i = 1, \ldots, k)$ and this in the sense of average values $E_i^{(0)}$. If the null hypothesis were, for example, the uniform distribution we had $\varepsilon_{i0} = n/k \; \forall \; i = 1, \ldots, k$. The cumulative test statistic $X^2(n)$ converges to the χ^2 distribution in the limit $n \to \infty$ – just as the average value of a stochastic variable, $\langle \mathcal{Z} \rangle \, = \, \sum_{i=1}^{n} z_i/n$ converges to the expectation value $\lim_{n \to \infty} \langle \mathcal{Z} \rangle \, = \, \mathrm{E}(\mathcal{Z})$. This implies that $X^2(n)$ is never exactly equal to χ^2 and the approximation that will always become better when the sample size is increased. Empirical knowledge of statisticians defines a lower limit for the number of entries in the cells to be considered, which lies between 5 and 10.

If the null hypotheses H_0 were true, ν_i and ε_{i0} should be approximately equal. Thus we expect the deviation expressed by

$$X_d^2 \; = \; \sum_{i=1}^{k} \frac{(\nu_i \, - \, \varepsilon_{i0})^2}{\varepsilon_{i0}} \; \approx \; \chi_d^2 \tag{2.114}$$

should be small if H_0 is acceptable. If the deviation is too large, we shall reject $H_0 : X_d^2 \geq \chi_d^2(\alpha)$, where α is the predefined level of significance for the test. Two basic quantities are still undefined (i) the degree of freedom d and (ii) the significance level α.

First the number of degrees of freedom d of the theoretical distribution to which the data are fitted has to be determined. The number of cells, k, represents the maximal number of degrees of freedom, which is reduced by one because of the conservation relation, $\sum_i \nu_i = n$, discussed above: $d = k - 1$. The dimension d is reduced further when parameters are needed in fitting the distribution of the null hypothesis. If the number of such parameters is s we get $d = k - 1 - s$. Choosing the parameter free uniform distribution $\mathcal{U}$ as null hypothesis we find, of course, $d = k - 1$.

The significance of the null hypothesis for a given set of data is commonly tested by means of the so-called p-value: For $p < \alpha$ the null hypothesis is rejected. Precisely, the p-value is the probability of obtaining a test statistic, which is at least as extreme as the actually observed one under the assumption that the null hypothesis is true. We call a probability $P(A)$ *more extreme* than $P(B)$ if A is less likely to occur than B under the null hypothesis. As shown in figure 2.23 this probability is obtained as the integral below the probability density function from the calculated X_d^2-value to $+\infty$. For the χ_d^2 distribution we have

$$p \;=\; \int_{X_d^2}^{+\infty} \chi_d^2(x)\,\mathrm{d}x \;=\; 1 \;-\; \int_0^{X_d^2} \chi_d^2(x)\,\mathrm{d}x \;= 1 \;-\; F(X^2;d)\;, \qquad (2.115)$$

which involves the cumulative distribution function of the χ^2-distribution, $F_{\chi^2}(x;d)$, defined in equation (2.65). Commonly, the null hypothesis is rejected when p is smaller than the significance level: $p < \alpha$ with $0.02 \leq \alpha \leq 0.05$. If the condition $p < \alpha$ is fulfilled one says the null hypothesis is rejected by statistical significance. In other words, the null hypothesis is statistically significant or statistically confirmed in the range $\alpha \leq p \leq 1$.

A simple example is used for the purpose of illustration: Two random samples of n animals were drawn from a population, ν_1 were males and ν_2 were females with $\nu_1 + \nu_2 = n$. The first sample $\nu_1 = 170, \nu_2 = 152$,

$$n = 322,\, \nu_1 = 170,\, \nu_2 = 152 : X_1^2 = \frac{(170-161)^2 + (152-161)^2}{322} = 0.503\,,$$
$$p = 1 - F_{\chi^2}(0.503;1) = 0.478\,,$$

clearly supports the null hypothesis that that males and females are equally frequent since $p > \alpha \approx 0.05$. The second sample $\nu_1 = 207, \nu_2 = 260$,

$$n = 467,\, \nu_1 = 207,\, \nu_2 = 260 : X_1^2 = \frac{(207-233.5)^2 + (260-233.5)^2}{233.5} = 6.015\,,$$
$$p = 1 - F_{\chi^2}(6.015;1) = 0.0142\,,$$

leads to a p-value, which is below the critical limit of significance and the rejection of the null hypothesis, the numbers of males and females are equal, is statistically significant or there is very likely another reason than random fluctuation responsible for the difference.

As a second example we test Gregor Mendel's experimental data on the garden pea, *Pisum sativum*, given in table 1.1. Here the null hypothesis to be tested is the ratio between different phenotypic features developed by the genotypes. We consider two features: (i) the shape, roundish and wrinkled, and (ii) the color of seeds, yellow and green, which are determined by two independent loci and two alleles each, A and a or B and b, respectively. The two alleles form four diploid genotypes, AA, Aa, and aA, aa, or, BB, Bb, and bB, bb, respectively. Since the alleles a and b are *recessive* only the the genotypes aa or bb develop the second phenotype, wrinkled and green, respectively, and based on the null hypothesis of a uniform distribution of genotypes we expect a 3:1 ratio of phenotypes. In table 2.3 we apply Pearson's chi-square hypothesis to the null hypothesis of 3:1 ratios for the phenotypes roundish and wrinkled or yellow and green. As examples we have chosen the total sample of Mendel's experiments as well as three plants ('1', '5', and '8') in table 1.1) being typical ('1') or showing extreme ratios ('5' having the best and the worst value for shape and color, respectively, and '8' showing the

Table 2.3 Pearson χ^2-test of Gregor Mendel's experiments with the garden pea (*Pisum sativum*)**.** The total results as well as the data for three selected plants are analyzed by means of Karl Pearson's chi-square statistics. Two characteristic features of the seeds are reported: the shape, roundish or angular wrinkled, and the color, yellow or green. The phenotypes of the two dominant alleles are: A = *round* and B = *yellow* and the recessive phenotypes are a = *wrinkled* and b = *green*. The data are taken from table 1.1.

Property	Sample space	Number of seeds		χ^2-statistics	
		A/B	a/b	X_1^2	p
shape (A,a)	total	5 474	1 850	0.2629	0.6081
color (B,b)	total	6 022	2 001	0.0150	0.9025
shape (A,a)	plant 1	45	12	0.4737	0.4913
color (B,b)	plant 1	25	11	0.5926	0.4414
shape (A,a)	plant 5	32	11	0.00775	0.9298
color (B,b)	plant 5	24	13	2.0405	0.1532
shape (A,a)	plant 8	22	10	0.6667	0.4142
color (B,b)	plant 8	44	9	1.8176	0.1776

largest ratio, 4.89). All p-values in this table are well above the critical limit and without further discussion required confirm the 3:1 ratio.[28]

The test of independence is relevant for situations when an observer registers two outcomes and the null hypothesis is that these outcomes are statistically independent. Each observation is allocated to one cell of a two-dimensional array of cells called a *contingency table* (see next section 2.6.3). In the general case there are m rows and n columns in a table. Then, the theoretical frequency for a cell under the null hypothesis of independence is

$$\varepsilon_{ij} \;=\; \frac{\sum_{k=1}^{n} \nu_{ik} \sum_{k=1}^{m} \nu_{kj}}{N} \;, \tag{2.116}$$

where N is the (grand) total sample size or the sum of all cells in the table. The value of the X^2 test-statistic is

$$X^2 \;=\; \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(\nu_{ij} - \varepsilon_{ij})^2}{\varepsilon_{ij}} \;. \tag{2.117}$$

Fitting the model of independence reduces the number of degrees of freedom by $\pi = m+n-1$. Originally the number of degrees of freedom is equal to the

[28] We should remember that the claim in a nutshell of Ronald Fisher and others had been that Mendel's data are too good to be true.

number of cells, $m \cdot n$, and after reduction by π we have $d = (m-1) \cdot (n-1)$ degrees of freedom for comparison with the χ^2 distribution. The p-value is again obtained by insertion into the cumulative distribution function (cdf), $p = 1 - F_{\chi^2}(X^2; d)$, and a value of p less than a predefined critical value, commonly $p < \alpha = 0.05$, is considered as justification for rejection of the null hypothesis or in other words the row variable does not appear to be independent of the column variable.

2.6.3 Fisher's exact test

As a second example out of many statistical significance test developed in mathematical statistics we mention here *Fisher's exact test* for the analysis of contingency tables. In contrast to the χ^2-test Fisher's test is valid for all sample sizes and not only for sufficiently large samples. We begin by defining a contingency table, which is a $m \times n$ matrix M where all possible outcomes of one variable x enter different columns in a row defined by a given outcome for y, and alternatively the distribution of outcomes of the second variable y for a specified outcome of x is contained in a column. The most common case – and the one that is most easily analyzed – is 2×2, two variables with two values each. The the contingency table has the form

	x_1	x_2	total
y_1	a	b	$a+b$
y_2	c	d	$c+d$
total	$a+c$	$b+d$	N

where every variable, x and y, has two outcomes and $N = a+b+c+d$ is the grand total. Fisher's contribution was to prove that the probability to obtain the set of values (x_1, x_2, y_1, y_2) is given by the hypergeometric distribution

$$\begin{aligned} \text{probability mass function} \quad & f_{\mu,\nu}(k) = \frac{\binom{\mu}{k}\binom{N-\mu}{\nu-k}}{\binom{N}{\nu}} \ , \\ \text{cumulative density function} \quad & F_{\mu,\nu}(k) = \sum_{i=0}^{k} \frac{\binom{\mu}{k}\binom{N-\mu}{\nu-k}}{\binom{N}{\nu}} \ , \end{aligned} \tag{2.118}$$

where $N \in \mathbb{N} = \{1, 2, \ldots\}$, $\mu \in \{0, 1, \ldots, N\}$, $\nu \in \{1, 2, \ldots, N\}$, and the support $k \in \{\max(0, \nu+\mu-N), \ldots, \min(\mu, \nu)\}$. Translating the contingency table into the notation of probability functions we have: $a \equiv k$, $b \equiv \mu - k$, $c \equiv \nu - k$, and $d \equiv N + k - (\mu + \nu)$ and hence Fisher's result for the p-value of the general 2×2 contingency table is

$$p = \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{N}{a+c}} = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!\,b!\,c!\,d!\,N!}\,, \tag{2.119}$$

where the expression on the rhs shows beautifully the equivalence between rows and columns.

We present the right- or left-handedness of human males or females as an example for the illustration of Fisher's test: A sample consisting of 52 males and 48 females yields 9 left-handed males and 4 left-handed females. Is the difference statistically significant and allows for the conclusion that left-handedness is more common among males than females? The contingency table in this case reads

	x_m	x_f	total
y_r	43	44	87
y_l	9	4	13
total	52	48	100

The calculation yields $p \approx 0.10$, which is above the critical value $0.02 \leq \alpha \leq 0.05$, and $p > \alpha$ confirms the null hypothesis of men and women being equally likely to be left-handed. Therefore, the assumption that males are more likely to be left-handed can be rejected for this data sample.

2.6.4 Bayesian inference

In this section we present a simple but analytically tractable example as an illustration for the application of Bayesian statistics [98], which has been adapted from the original work of Reverend Thomas Bayes in the posthumous publication in 1763 [375]. More detailed applications of the Bayesian approach can be found in a number of excellent monographs, for example in [159].

The example is called table game and is played by two persons, Alice (A) and Bob (B) as well as a third person (C) acting as game master and being *neutral*. A (pseudo)random number generator is used to draw pseudorandom numbers from a uniform distribution in the range $0 \leq \mathcal{R} < 1$. The pseudorandom number generator is operated by the game master and cannot be seen by the two players. In essence, A and B are completely passive, they have no information on the game except knowledge on the basic setup of the game and they know the scores, which are $a(t)$ for A and $b(t)$ for B. The person who reaches a predefined score value, z, first has won. This simple game starts through drawing a pseudorandom number, $\mathcal{R} = r_0$, by the game master. Consecutive drawings yielding numbers r_i assign points to A iff $0 \leq r_i < r_0$ is fulfilled and to B iff $r_0 \leq r_i < 1$ holds. The game is continued until one person, A or B, reaches the score z.

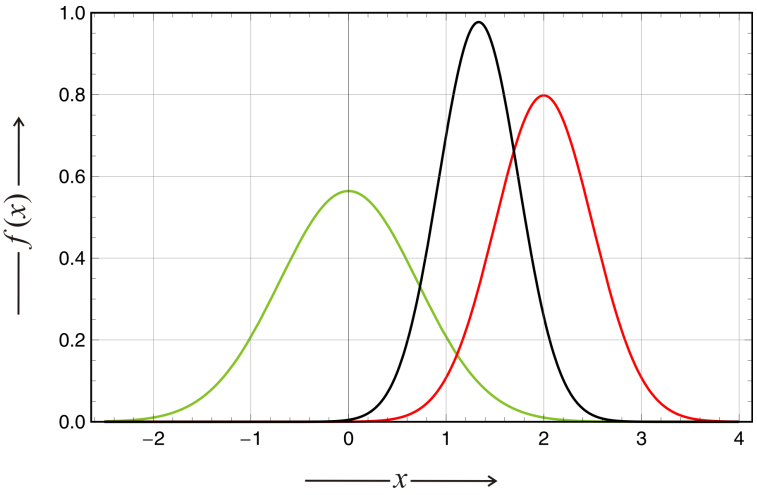


Fig. 2.25 The Bayesian method of inference. The figure sketches the Bayesian method by means of normal density functions. The sample data are given in form of the likelihood function ($P(\mathcal{Y}|\mathcal{X}) = \mathcal{N}(2, \frac{1}{2})$, red) and additional external information on the parameters enters the analysis as prior distribution ($P(\mathcal{X}) = \mathcal{N}(0, 1/\sqrt{2})$, green). The resulting posterior distribution $P(\mathcal{X}|\mathcal{Y}) = P(\mathcal{Y}|\mathcal{X}) \cdot P(\mathcal{X})/P(\mathcal{Y})$ (black) is here again a normal distribution with mean $\bar{\mu} = (\mu_1\sigma_2^2 + \mu_2\sigma_1^2)/(\sigma_1^2 + \sigma_2^2)$ and variance $\bar{\sigma}^2 = (\sigma_1^2\sigma_2^2)/(\sigma_1^2 + \sigma_2^2)$. It is straightforward to show that the mean $\bar{\mu}$ lies between μ_1 and μ_2 and variance has become smaller $\bar{\sigma} \leq \min(\sigma_1, \sigma_2)$ (see text).

The problem is to compute fair odds of winning for A and B when the game is terminated premature, and r_0 is unknown. Let us assume that the scores at the time of termination were: $a(t) = a$ and $b(t) = b$ with $a < z$ and $b < z$, and to make the calculations easy we assume that A is only one point away from winning, $a = z - 1$ and $b < z - 1$. If r_0 were known the answer would be trivial. In the conventional approach we would make an assumption about the parameter r_0. In the lack of knowledge we could make the null hypothesis $r_0 = \hat{r}_0 = \frac{1}{2}$, and find simply

$$P_0(B) \;=\; P(B \text{ is winning}) \;=\; (1-\hat{r}_0)^{z-b} \;=\; \left(\frac{1}{2}\right)^{z-b} \;,$$

$$P_0(A) \;=\; P(A \text{ is winning}) \;=\; 1 \,-\, (1-\hat{r}_0)^{z-b} \;=\; 1 \,-\, \left(\frac{1}{2}\right)^{z-b} \;,$$

because the only way for B to win is to make $z - b$ scores in a row. Thus fair odds for A to win would be $(2^{z-b} - 1) : 1$. An alternative approach is to make the *maximum likelihood* estimate on the unknown parameter $r_0 = \tilde{r}_0 = a/(a+b)$ and again we calculate the probabilities and find by the same token

$$P_{\mathrm{ml}}(B) \;=\; P(B \text{ is winning}) \;=\; (1-\tilde{r}_0)^{z-b} \;=\; \left(\frac{b}{a+b}\right)^{z-b} \;,$$

$$P_{\mathrm{ml}}(A) \;=\; P(A \text{ is winning}) \;=\; 1 \,-\, (1-\tilde{r}_0)^{z-b} \;=\; 1 \,-\, \left(\frac{b}{a+b}\right)^{z-b} \;,$$

and for the odds in favor of A: $\left(\frac{a+b}{b}\right)^{z-b} - 1$.

The Bayesian solution considers $r_0 = p$ as a unknown but variable parameter about which no estimate is made. Instead the uncertainty is modeled rigorously by integrating over all possible values: $0 \le p \le 1$. The expected probability for B to win is then

$$\mathrm{E}\Big(P(B)\Big) \;=\; \int_0^1 (1-p)^{z-b}\, P(p\,|a,b)\,\mathrm{d}p \;,$$

where $(1-p)^{z-b}$ is the probability for winning of B and $P(p\,|a,b)$ is the probability of a certain value of p provided the data a and b were obtained at the termination of the game. The probability $P(p\,|a,b)$ formally written as $P(\text{model}|\text{data})$ is the inversion of the common problem $P(\text{data}|\text{model})$ – given a certain model what is the probability to find a certain set of data – and a so-called *inverse probability* problem. The solution of the problem is provided by Bayes' theorem, which is an almost trivial truism for two random variables $\mathcal{X}$ and $\mathcal{Y}$:

$$P(\mathcal{X}|\mathcal{Y}) \;=\; \frac{P(\mathcal{Y}|\mathcal{X})\cdot P(\mathcal{X})}{P(\mathcal{Y})} \;=\; \frac{P(\mathcal{Y}|\mathcal{X})\cdot P(\mathcal{X})}{\sum_{\mathcal{Z}} P(\mathcal{Y}|\mathcal{Z})\cdot P(\mathcal{Z})} \;, \qquad (1.4')$$

where the sum over the random variable $\mathcal{Z}$ covers entire sample space. Equation (1.4') yields in our example

$$P(p\,|a,b) \;=\; \frac{P(a,b\,|p)\cdot P(p)}{\int_0^1 P(a,b\,|\varrho)\cdot P(\varrho)\,\mathrm{d}\varrho} \;.$$

The interpretation of the equation is straightforward: The probability of a particular choice of p given the data (a,b) called the *posterior probability* (figure 1.3) is proportional to the probability to obtain the observed data if p were true – the *likelihood* of p – multiplied by the *prior probability* of this particular value of p relative to all other values of p. The integral in the denominator takes care of the normalization of the probability – the summation is replaced by an integral, because p is a continuous variable, and $0 \le p \le 1$ is the entire domain of p.

The likelihood term is calculated readily from the binomial distribution

$$P(a,b\,|p\,) \;=\; \binom{a+b}{b} p^a\,(1-p)^b \;,$$

but the probability prior requires more care. By definition $P(p)$ is the probability of p before the data have been recorded. How can we estimate p before we have seen any data? We are thus referred to the situation how r_0 is determined, and we know it has been picked from the uniform distribution and hence, $P(p)$ is a constant that appears in the numerator and in the denominator and thus cancels in the equation for Bayes' theorem (1.4'). After some algebraic computation we eventually obtain for winning of B:

$$\mathrm{E}\Big(P(B)\Big) \;=\; \frac{\int_0^1 p^{\,a}\,(1-p)^z\,\mathrm{d}p}{\int_0^1 p^{\,a}\,(1-p)^b\,\mathrm{d}p}\;.$$

Integration is straightforward, because the integrals are known as Euler integrals of the first kind, which have the Beta-function as solution

$$B(x,y) \;=\; \int_0^1 z^{x-1}\,(1-z)^{y-1}\,\mathrm{d}z \;=\; \frac{(x-1)!\,(y-1)!}{(x+y-1)!} \;=\; \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}\,. \qquad (2.120)$$

Finally, we obtain the following expression for the chance of winning of B

$$\mathrm{E}\big(P(B)\big) \;=\; \frac{z!\,(a+b+1)!}{b!\,(a+z+1)!}\;,$$

and the Bayesian estimation for fair odds yields

$$\left(\frac{b!\,(a+z+1)!}{z!\,(a+b+1)!} - 1\right) \;:\; 1\;.$$

A specific numerical example is given in [98]: $a = 5$, $b = 3$, and $z = 6$. The null hypothesis of equal probabilities of winning for A and B, $\hat{r}_0 = 0.5$ yields an advantage of 7:1 for A, the maximum likelihood approach with $\tilde{r}_0 = a/(a+b) = 5/8$ yields ≈18:1, and the Bayesian estimate yields 10:1. The large differences should not be surprising since the sample size is very small. The correct answer of the table game with the values for a, b, and z is indeed 10 as can be easily verified by numerical computation with a small computer program.

Finally, we show how the Bayesian approach operates on probability distributions (a simple but straightforward description is found in [415]). According to equation (1.4') the posterior probability $P(\mathcal{X}|\mathcal{Y})$ is obtained through multiplication of the prior probability $P(X)$ by the data likelihood function $P(\mathcal{Y}|\mathcal{X})$ and normalization. We illustrate the relation between the probability function by means of two normal distributions and their product (figure 2.25). For the prior probability and the data function we assume

$$P(\mathcal{X}) \;=\; f_1(x) \;=\; \frac{1}{\sqrt{2\pi\sigma_1^2}}\, e^{-(x-\mu_1)^2/(2\sigma_1^2)} \;\text{ and}$$

$$P(\mathcal{X}|\mathcal{Y}) \;=\; f_2(x) \;=\; \frac{1}{\sqrt{2\pi\sigma_2^2}}\, e^{-(x-\mu_2)^2/(2\sigma_2^2)}\;,$$

and obtain for the product with the normalization factor $\mathcal{N} = \mathcal{N}(\mu_1, \mu_2, \sigma_1, \sigma_2)$

$$P(\mathcal{Y}|\mathcal{X}) \;=\; \mathcal{N}\, f_1(x)\, f_2(x) \;=\; \mathcal{N}\, g\; e^{-(x-\bar{\mu})^2/(2\bar{\sigma}^2)} \;\text{ with}$$

$$\bar{\mu} \;=\; \frac{\mu_1\sigma_2^2 + \mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}\,,\quad \bar{\sigma}^2 \;=\; \frac{\sigma_1^2\,\sigma_2^2}{\sigma_1^2 \,+\, \sigma_2^2}\,,\quad g \;=\; \frac{1}{2\pi\,\sigma_1\,\sigma_2}\, e^{-\frac{1}{2}\frac{(\mu_2-\mu_1)^2}{\sigma_1^2+\sigma_2^2}}\;,\;\text{ and}$$

$$\mathcal{N}\, g \;=\; \frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sqrt{2\pi}\;\sigma_1\,\sigma_2} \;=\; \frac{1}{\sqrt{2\pi\,\bar{\sigma}^2}}\,,$$

as required for normalization of the Gaussian curve.

Two properties of the posterior probability are easily tested by means of our example: (i) the averaged mean, $\bar{\mu}$, lies always between μ_1 and μ_2 and (ii) the product distribution is sharper than the two factor distributions

$$\frac{\sigma_1^2\,\sigma_2^2}{\sigma_1^2+\sigma_2^2} \;\leq\; \min\{\sigma_1^2, \sigma_2^2\}\;,$$

with the equals sign requiring either $\sigma_1 = 0$ or $\sigma_2 = 0$. The improvement of the Bayesian analysis thus reduces the difference in the mean values between expectation and model, and the distribution becomes narrower in the sense of reducing uncertainty.

Whereas the Bayesian approach does not seem to provide a lot more information in situations where the models are confirmed by many other independent applications like, for example, in the majority of problems in physics and chemistry, the highly complex situations in modern biology, economics, or social sciences require highly simplified and flexible models and there is an ample field for application of Bayesian statistics.

Chapter 3
Stochastic processes

With four parameters I can fit an elephant and with five I can make him wiggle his trunk.
Enrico Fermi citing John von Neumann, 1953 [97].

Abstract Stochastic processes are defined and grouped into different classes, their basic properties are listed and compared. The Chapman-Kolmogorov equation is introduced, transformed into a differential version, and used for the classification of the major types of processes: (i) drift and (ii) diffusion with continuous sample paths, and (iii) jump processes being essentially discontinuous. In pure form these prototypes are described by Liouville equations, stochastic diffusion equations, and master equations, respectively. The most popular and most frequently used continuous equation is the Fokker-Planck (FP) equation that describes the evolution of a probability density by drift and diffusion. The pendant to FP equations on the discontinuous side are master equations, which are dealing with jump processes only and represent the appropriate tool for modeling processes described by discrete variables. For technical reasons they are often difficult to handle unless population sizes are relatively small. Stochastic differential equations (SDEs) model processes at the level of random variables by solving ordinary differential equations upon which a diffusion process, called a Wiener process, is superimposed. Ensembles of individual trajectories of SDEs are equivalent to time dependent probability densities described by Fokker-Planck equations.

Stochastic processes introduce time into probability theory and represent the most prominent possibility to combine dynamical phenomena and randomness resulting from incomplete information. In physics and chemistry the dominant source of randomness is thermal motion but in biology the overwhelming complexity of systems is commonly prohibitive for a complete description and then, lack of information results also from simplifications inherent in the model. In essence, there are two ways of dealing with stochasticity in processes: (i) calculation or recording of stochastic variables as functions of time and (ii) modeling of the temporal evolution of entire probability densities. In case (i) one particular computation or experiment yields a single sample path or a *trajectory*, and full information on the process is obtained

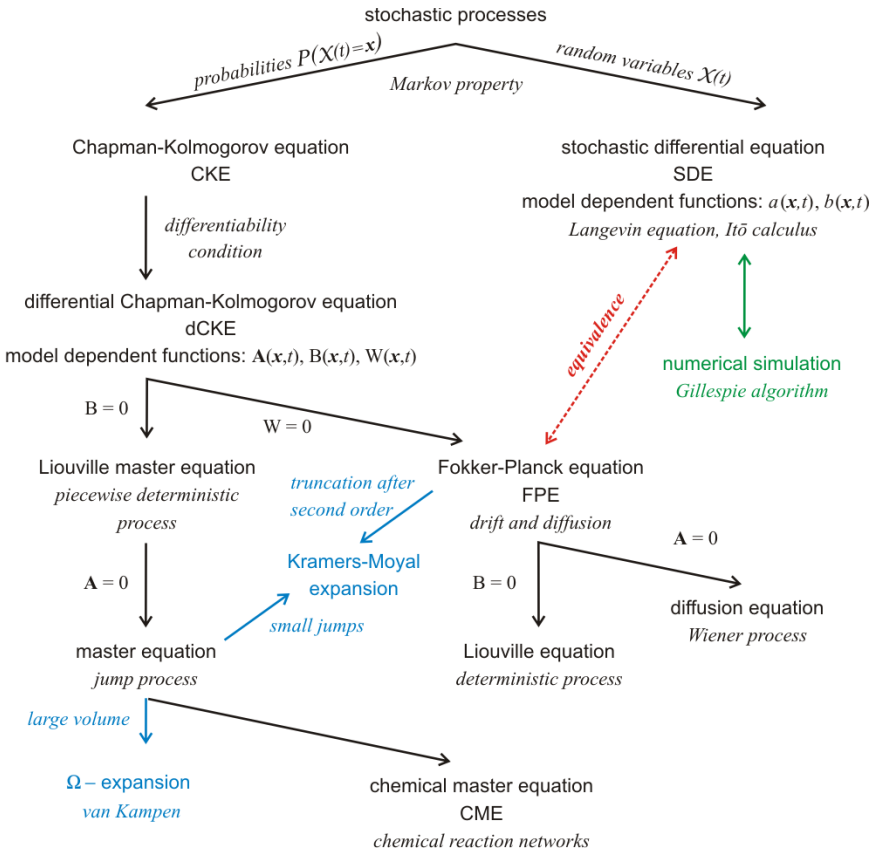


Fig. 3.1 Description of stochastic processes. The sketch presents a *family tree* of stochastic models [434]. Almost all stochastic models used in science are based on the Markov property of processes, which – in a nutshell – states that full information on the system at present is sufficient for predicting or modeling the future (section 3.1.3.3). Models fall into two major classes depending on the objects they are dealing with: (i) random variables $\mathcal{X}(t)$ or (ii) probability densities $P\big(\mathcal{X}(t) = x\big)$. In the center of stochastic modeling stands the Chapman-Kolmogorov equation (CKE) that introduces the Markov property into time series of probability densities. In differential form CKE contains three model dependent functions, the vector $\mathbf{A}(\mathbf{x}, t)$ and the matrices $\mathrm{B}(\mathbf{x}, t)$, and $\mathrm{W}(\mathbf{x}, t)$, which determine the nature of the stochastic process. Different combinations of these functions yield the most important equations for stochastic modeling: the Fokker-Planck equation with $\mathrm{W} = 0$ ($\mathbf{A} \neq 0$ and $\mathrm{B} \neq 0$), the stochastic diffusion equation with $\mathrm{B} \neq 0$ ($\mathbf{A} = 0$ and $\mathrm{W} = 0$), and the master equation with $\mathrm{W} \neq 0$ ($\mathbf{A} = 0$ and $\mathrm{B} = 0$). For stochastic processes without jumps the solutions of the stochastic differential equation are trajectories, which when properly sampled describe the evolution of a probability density $P\big(\mathcal{X}(t) = x\big)$ that is equivalent to the solution of a Fokker-Planck equation (red arrow). Common approximations by means of size expansions are shown in blue. The green arrow indicates where conventional numerical simulations come into play.

by sampling of trajectories from repetitions under identical conditions.[1] Sampling of trajectories leads to bundles of curves, which can be evaluated in the spirit of mathematical statistics (section 2.6) to yield time dependent moments of the probability densities. For an illustrative example comparing superposition of trajectories and migration of the probability density we refer to the Ornstein-Uhlenbeck process shown in figures 3.9 and 3.10.

The expectation value of a random variable as a function of time, $\mathrm{E}\big(\mathcal{X}(t)\big)$, often coincides with the deterministic solution of the corresponding differential equation. For single point initial conditions the solution curves of ordinary or partial differential equations (ODEs or PDEs) consists of single trajectories as determined by the theorems of existence and uniqueness of solutions. Solutions of stochastic processes as said above correspond to bundles of trajectories, which differ in the sequence of random events and which as a rule surround the deterministic solution. Commonly, sharp initial condition are chosen and then the bundle of trajectories starts in a single point and diverges into the future as well as into the past depending on whether the process is studied in the *forward* or in the *backward* direction (see figure 3.22). The stochastic equations in forward and backward directions are different and the typical symmetry of differential equations with respect to time reversal is no more existent because of the diffusion term [7, 110, 408]. In the forward direction the time dependent variance, $\mathrm{var}\big(\mathcal{X}(t)\big)$ allows for a useful distinction of two types of stochastic processes: (i) The variance increases with time and grows without limits, a behavior that is typical for unlimited spatial diffusion and some biologically important processes involving populations in abstract spaces, and (ii) the variance approaches a finite long time limit, which corresponds to the thermodynamic equilibrium or to a stationary state and where the standard deviation fulfils an approximate $\sqrt{N}$-law.

Figure 3.1 presents an overview of the most frequently used general model equations for stochastic processes,[2] which are introduced in this chapter, and it shows how they are interrelated [434, 435]. Two classes of equations are of central importance: (i) the differential form of the Chapman-Kolmogorov equation (dCKE; section 3.2) describing the evolution of probability densities and (ii) the stochastic differential equation (SDE; section 3.4) modeling stochastic trajectories. The Fokker-Planck equation and the master equation are derived from the differential Chapman-Kolmogorov equation through restriction to continuous processes or jump processes, respectively. The *chemical master equation* is a master equation adapted for modeling chemical reaction networks where the jumps are changes in the integer particle numbers of chemical species (section 4.2.1). In this chapter we shall present a brief introduction into stochastic processes and into the general formalisms for modeling

[1] Identical conditions means that all parameters are the same except the random fluctuations. In computer simulations this is achieved by keeping everything the same except the seeds for the pseudorandom number generator.

[2] By *general* we mean here methods that are widely applicable and not tailored specifically for deriving solutions for one case or a few cases only.

them. The chapter is essentially based on three textbooks [76, 157, 441] and it uses in essence the notation introduced by Crispin Gardiner [156]. A few examples of stochastic processes of general importance will be discussed here for the purpose of illustration of the formalisms. In particular we shall focus on random walks and diffusion. Other applications are presented in the forthcoming two chapters 4 and 5. Analysis of stochastic processes by mathematics is complemented by numerical simulations [173], which became more and more important over the years essentially for two reasons: (i) the accessibility of cheap and extensive computing power, and (ii) the need for stochastic treatments of complex kinetics in chemistry and biology that escapes analytical handling. Numerical simulation methods will also be presented and discussed here in chapter 4 (section 4.6).

3.1 Modeling stochastic processes

The use of conventional differential equations for modeling dynamical systems implies determinism in the sense that full information about the system at a single time, t_0 for example, allows for exact computation of future and past. In reality we encounter substantial limitations concerning prediction and reconstruction, especially in case of deterministic chaos because initial and boundary conditions are available only with finite accuracy, and even the smallest errors are amplified to any size after sufficiently long time. The theory of *stochastic processes* provides the tools for taking into account all possible sources of uncontrollable irregularities, and defines in a natural way the limits for predictions of the future as well as reconstruction of the past. Different stochastic processes can be classified with respect to *memory effects* making precise how the past acts on the future. Almost all stochastic models in science fall into the very wide class of *Markov processes*, which are named after the Russian mathematician Andrey Markov,[3] and which are characterized the lack of memory in the sense that the future can be modeled and predicted probabilistically form knowledge of the presence and no information on historical events is required.

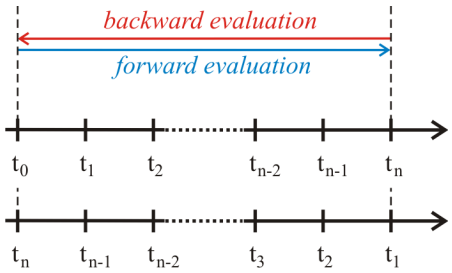


Fig. 3.2 Time order in modeling stochastic processes. Physical or real time is progressing from left to right and the most recent event is given by the rightmost recording. Conventional numbering of instances in physics starts at some time t_0 and ends at time t_n (upper time axis). In the theory of stochastic processes an opposite ordering of times is often preferred and then t_1 is the latest event of the series (lower time axis). Modeling stochastic processes, for example by a Chapman-Kolmogorov equation, distinguishes two modes of description: (i) the forward equation predicting the future from past and present and (ii) the backward equation that extrapolates back in time from present to past. Accordingly, we have a second time scale used in the computation, which progresses in the same direction as real time for the forward evaluation (blue) and in opposite direction for the backward evaluation (red).

[3] The Russian mathematician Andrey Markov (1856-1922) is one of the founders of Russian probability theory and pioneered the concept of memory free processes, which are named after him. He expressed more precisely the assumptions that were made by Albert Einstein [108] and Marian von Smoluchowski [453] in their derivation of the diffusion process.

3.1.1 Trajectories and processes

The probabilistic evolution of a system in time is described as a general stochastic process. We assume the existence of a time dependent random variable $\mathcal{X}(t)$ or random vector $\vec{\mathcal{X}}(t) = \big(\mathcal{X}_k(t); k = 1, \ldots, M; k \in \mathbb{N}_{>0}\big)$.[4] As in the previous chapters we shall distinguish the simpler discrete case,

$$P_n(t) \;=\; P\Big(\mathcal{X}(t) = x_n\Big) \;\text{ with }\; n \in \mathbb{N}\,, \tag{3.1}$$

from the continuous or probability density case,

$$dF(x,t) \;=\; f(x,t)\,dx \;=\; P\Big(x \le \mathcal{X}(t) \le x + dx\Big) \;\text{ with }\; x \in \mathbb{R}\,. \tag{3.2}$$

It is easier to visualize the discrete case first and postpone the generalization to densities, which as such is straightforward. A particular series of events – be it the result of a calculation or an experiment – constitutes a *sample path* or a *trajectory* in phase space.[5] The trajectory consists of a list of the recorded values of the random variable $\mathcal{X}$ at certain times arranged in the form of doubles (x_i, t_i)

$$\mathcal{T} \;=\; \Big((x_1,t_1),(x_2,t_2),(x_3,t_3),\cdots,(x_k,t_k),(x_{k+1},t_{k+1}),\cdots\Big)\,. \tag{3.3}$$

Although it is not essential for the application of probability theory, but for the sake of clearness we shall always assume that the recorded values are time ordered, here with the earliest or oldest values on the rightmost position and the most recent values at the latest entry on the left-hand side. Assuming that the recorded series has started at some time t_n in the past with x_n we have

$$t_1 \ge t_2 \ge t_3 \ge \cdots \ge t_k \ge t_{k+1} \ge \cdots\,.$$

Accordingly a trajectory is a sequence of time ordered doubles (x, t).

It is worth noticing that the conventional way of counting time in physics progresses in opposite direction from some initial time $t = t_0$ to t_1, t_2, t_3 and so on until t_n, the most recent instant is reached (figure 3.2):

---

[4] For the moment we need not specify whether $\mathcal{X}(t)$ is a simple random variable or a random vector $\vec{\mathcal{X}}(t) = \big(\mathcal{X}_k(t); k = 1, \ldots, M\big)$ and therefore we drop the index 'k' determining the individual component. Later on, for example in chemical kinetics, when the distinction of different (chemical) species becomes necessary, we shall make clear in which sense $\mathcal{X}(t)$ is used, random variable or random vector.

[5] Here we shall use the notion of *phase space* in a loose way for an abstract space that is sufficient for the characterization of the system and for the description of its temporal development. For example in a reaction involving n chemical species the phase space will be a Cartesian space spanned by n axes for n concentrations. In classical mechanics the phase space is precisely defined as a – commonly Cartesian – space spanned by the $3n$ spatial coordinates and the $3n$ coordinates of the linear momenta of an n particle system, and the same definition is used in statistical mechanics.

$$\mathcal{T} \;=\; \Big((x_n,t_n),(x_{n-1},t_{n-1}),\cdots,(x_k,t_k),(x_{k-1},t_{k-1}),\cdots,(x_0,t_0)\Big)\;, \quad (3.3')$$

where we adopt the same notation as in equation (3.3) with changed ordering

$$t_n \geq t_{n-1} \geq \cdots \geq t_k \geq t_{k-1} \geq \cdots \geq t_0\;.$$

In order to avoid confusion we shall always state explicitly when we are not using the convention shown in equation (3.3).[6]

Single trajectories are superimposed to yield bundles of trajectories in the sense of a summation of random variables as in equation (1.23):[7]

$$\begin{array}{cccc} \mathcal{X}^{(1)}(t_0) & \mathcal{X}^{(1)}(t_1) & \ldots & \mathcal{X}^{(1)}(t_n) \\ \mathcal{X}^{(2)}(t_0) & \mathcal{X}^{(2)}(t_1) & \ldots & \mathcal{X}^{(2)}(t_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}^{(N)}(t_0) & \mathcal{X}^{(N)}(t_1) & \ldots & \mathcal{X}^{(N)}(t_n) \\ \hline \mathcal{S}(t_0) & \mathcal{S}(t_1) & \ldots & \mathcal{S}(t_n) \end{array}\;,$$

and we obtain the summation random variable $\mathcal{S}(t)$. The calculation of sample moments is straightforward and from equations (2.101) and (2.104) we get:

$$\begin{aligned} m(t) \;&=\; \widetilde{\mu}(t) \;=\; \frac{1}{N}\,\mathcal{S}(t) \;=\; \frac{1}{N}\sum_{i=1}^{N} x^{(i)}(t)\,, \\ m_2(t) \;&=\; \widetilde{\mathrm{var}}(t) \;=\; \frac{1}{N-1}\sum_{i=1}^{N}\big(x^{(i)}(t)-m(t)\big)^2 \;= \\ &=\; \frac{1}{N-1}\left(\sum_{i=1}^{N} x^{(i)}(t)^2 \;-\; N\,m(t)^2\right)\,. \end{aligned} \quad (3.4)$$

An illustration by means of a numerical example is shown in figure 3.3.

In chapters 1 and 2 we have used the vague notion of *scores* and not yet specified, which quantities the random variables $(\mathcal{A},\mathcal{B},\ldots,\mathcal{W}) \in \Omega$ actually describe and what their realizations in some measurable space, $(a,b,\ldots,w) \in \mathbb{R}$, precisely are. So far almost all events and samples were expressed as dimensionless numbers. Considering processes introduces time and time has a dimension and accordingly we need to specify a unit in which the recorded data are measured – second, minute or hour, for example. Processes may take place in three-dimensional physical space where units for

[6] The different numberings for the elements of trajectories should not be confused with forward and backward processes (figure 3.2) to be discussed later on in section 3.3.

[7] In order to leave the subscript free for the indication of discrete times or different chemical species we use the somewhat clumsy superscript notation, $\mathcal{X}^{(i)}$ or $x^{(i)}$ $(i = 1,\ldots,N)$, to indicate individual trajectories, and we use the physical numbering of times, $t_0 \to t_n$.

Table 3.1 Notation used in modeling stochastic processes. Four different approaches to model stochastic processes by probability densities are compared: (i) discrete values of the random variable $\mathcal{X}$ and discrete time, (ii) discrete values and continuous time, (iii) continuous values and discrete time, and eventually (iv) continuous values and continuous time.

Values	Time: discrete	Time: continuous
discrete	$P_{n,k} = P(\mathcal{X}_k = x_n)\,;\, k,n \in \mathbb{N}$	$P_n(t) = \big(\mathcal{X}(t) = x_n\big)\,;\, n \in \mathbb{N}, t \in \mathbb{R}$
continuous	$p_k(x)\,\mathrm{d}x = P(x \le \mathcal{X}_k \le x + \mathrm{d}x) =$ $= f_k(x)\,\mathrm{d}x = \mathrm{d}F_k(x)$ $k \in \mathbb{N}\,,\ x \in \mathbb{R}$	$p\,(x,t)\,\mathrm{d}x = P(x \le \mathcal{X}_k \le x + \mathrm{d}x) =$ $= f(x,t)\,\mathrm{d}x = \mathrm{d}F(x,t)$ $x,t \in \mathbb{R}$

length, area and volume are required. In applications we shall be concerned with variables of other physical dimensions, for example mass, viscosity, surface tension, electric charge, magnetic moments, electromagnetic radiation, etc. Wherever a quantity is introduced we shall mention its dimension and the common units used in measurements.

Stochastic processes in chemistry and biology are commonly modeling the time development of ensembles or populations. In spatially homogeneous chemical reaction systems the variables are discrete particle numbers or continuous concentrations, $\mathcal{A}(t)$ or $a(t)$, and as a common notation we shall use $[\mathbf{A}(t)]$ and omit '(t)' whenever there is no misunderstanding possible. Spatial heterogeneity, for example, is accounted for by explicit consideration of diffusion and this leads to a reaction-diffusion systems, where the solutions can be visualized as migrations of evolving probability densities in time and three dimensional space. Then, the variables are functions in 3D space and time, $\mathcal{A}(\mathbf{r},t)$ or $a(\mathbf{r},t)$, with $\mathbf{r} = (x,y,z) \in \mathbb{R}^3$ being a vector in space. In biology the variables are often numbers of individuals in populations and then they depend on time or as in chemistry on time and on three-dimensional space when migration processes are considered. Sometimes it is of advantage to consider stochastic processes in formal spaces like the genotype or sequence space, which is a discrete space where the points represent individual genotypes and the distance of two genotypes commonly called Hamming-distance counts the minimal number of mutations required to bridge the interval between them. Neutral evolution, for example, can then be visualized as a diffusion process (section 5.3.2) and Darwinian selection as a hill climbing process in genotype space [469].

3.1.2 Probabilistic notation for processes

A stochastic process is determined by a set of joint probability densities the existence and analytical form of which is presupposed.[8] The probability density encapsulates the physical nature of the process and contains all parameters and data reflecting internal dynamics and external conditions and in this way determines completely the system under consideration:

$$p\,(x_1,t_1;x_2,t_2;x_3,t_3;\cdots;x_n,t_n;\cdots)\ . \tag{3.5}$$

By *complete determination* we mean that no additional information is required to describe the progress of the system as a time ordered series (3.3) and we shall call such a process a *separable stochastic process*. Although more general processes are conceivable, they play little role in current physics, chemistry, and biology and therefore we shall not consider them here.

Calculation of probabilities from (3.5) by means of marginal densities (1.40) and (1.72) is straightforward. For the discrete case the result is obvious:

$$P(\mathcal{X}=x_1)\ =\ p\,(x_1,*)\ =\ \sum_{x_k\neq x_1} p\,(x_1,t_1;x_2,t_2;x_3,t_3;\cdots;x_n,t_n;\cdots)\ .$$

The probability to record the value x_1 for the random variable $\mathcal{X}$ at time t_1 is obtained through summation over all previous values $x_2, x_3, \ldots$. In the continuous case the summations are simply replaced by integrals,

$$P(\mathcal{X}_1=x_1\in[a,b])\ =\ \int_a^b \mathrm{d}x_1 \iiint_{-\infty}^{\infty} \mathrm{d}x_2\mathrm{d}x_3\cdots\mathrm{d}x_n\cdots \\ p\,(x_1,t_1;x_2,t_2;x_3,t_3;\cdots;x_n,t_n;\cdots)\ .$$

Time ordering admits a formulation of the predictions of future values from the known past in terms of conditional probabilities:

$$p\,(x_1,t_1;x_2,t_2;\cdots|x_k,t_k;x_{k+1},t_{k+1},\cdots)\ =\ \frac{p\,(x_1,t_1;x_2,t_2;\cdots;x_k,t_k;x_{k+1},t_{k+1},\cdots)}{p\,(x_k,t_k;x_{k+1},t_{k+1},\cdots)}\ ,$$

with $t_1\geq t_2\geq\cdots\geq t_k\geq t_{k+1}\geq\cdots$. In other words, we may compute

$$\{(x_1,t_1),(x_2,t_2),\cdots\}\ \text{from known}\ \{(x_k,t_k),(x_{k+1},t_{k+1}),\cdots\}\ .$$

With respect to the temporal progress of the process we shall distinguish discrete and continuous time: A trajectory in discrete time is just a time ordered sequence of random variables, $\mathcal{X}_1,\mathcal{X}_2,\ldots,\mathcal{X}_n$ where time is implicitly included in the index of the variable in the sense that $\mathcal{X}_1$ is recorded at

[8] The joint density p is defined as in equations (1.37) and section 1.9.3 but with a slightly different notation. In stochastic processes we are always dealing with doubles (x,t), which we separate by a semicolon: $\cdots;x_k,t_k;x_{k+1},t_{k+1};\cdots$.

time t_1, $\mathcal{X}_2$ at time t_2, and so on. The discrete probability distribution is characterized by two indices, n for the integer values the random variable can adopt and k for time: $P_{n,k} = P(\mathcal{X}_k = x_n)$ with $n, k \in \mathbb{N}_{>0}$ (table 3.1). The introduction of continuous time is straightforward, since we need only replace $k \in \mathbb{N}_{>0}$ by $t \in \mathbb{R}$. The random variable is still discrete and the probability mass function becomes a function of time, $P_{n,k} \Rightarrow P_n(t)$. The transition to a continuous sample space for the random variable is done in precisely the same way as in the case of probability mass functions described in section 1.9. For the discrete time case we change the notation accordingly, $P_{n,k} \Rightarrow p_k(x)\mathrm{d}x = f_k(x)\mathrm{d}x = dF_k(x)$, and for continuous time we have $P_{n,k} \Rightarrow p\,(x,t)\mathrm{d}x = f(x,t)\mathrm{d}x = dF(x,t)\mathrm{d}x$.

Before we derive a general concept that allows for flexible modeling of stochastic processes, which are applicable to chemical kinetics and biological modeling, we introduce a few common classes of stochastic processes with certain characteristic properties that are meaningful in the context of applications. In addition we shall distinguish different behavior with respect to past, present and future encapsulated in memory effects.

3.1.3 Memory in stochastic processes

Three simple stochastic processes with characteristic memory effects will be discussed here: (i) the fully factorizable process with probability densities that are independent of other events with the special case of the Bernoulli process where the probability densities are also independent of time, (ii) the martingale where the (sharp) initial value of the stochastic variable is equal to the conditional mean value of the variable in the future, and (iii) the Markov process, where the future is completely determined by the presence and which is the most common formalism for modeling dynamics stochasticity in science.

3.1.3.1 Independence and Bernoulli processes

The simplest class of stochastic processes is characterized by *complete independence* of events that allows for factorization of the density,

$$p\,(x_1,t_1;x_2,t_2;x_3,t_3;\cdots) \;=\; \prod_i p\,(x_i,t_i)\;. \tag{3.6}$$

Equation (3.6) implies that the current value $\mathcal{X}(t)$ is completely independent of its values in the past. A special case is the sequence of Bernoulli trials (see in previous chapters, in particular in sections 1.5 and subsection 2.3.2) where the probability densities are also independent of time: $p\,(x_i,t_i) = p\,(x_i)$, and then we have

$$p\,(x_1,t_1;x_2,t_2;x_3,t_3;\cdots) \;=\; \prod_i p\,(x_i)\;. \tag{3.6'}$$

Further simplification occurs, of course, when all trials are based on the same probability distribution – for example, if the same coin is tossed in Bernoulli trials or the same dice are thrown – and then the product can be replaced by the power $p\,(x)^n$.

3.1.3.2 Martingales

The notion of *martingale* has been introduced by the French mathematician Paul Pierre Lévy and the development of the theory of martingales is due to the American mathematician Joseph Leo Doob [304]. Appropriately, we distinguish discrete time and continuous time processes. A discrete-time martingale is a sequence of random variables, $\mathcal{X}_1, \mathcal{X}_2, \ldots$, which satisfy the conditions[9]

$$\mathrm{E}(\mathcal{X}_{n+1}|\mathcal{X}_n, \ldots, \mathcal{X}_1) \;=\; \mathcal{X}_n \quad \text{and} \quad \mathrm{E}(|\mathcal{X}_n|) < \infty \;. \tag{3.7}$$

Given all past values $\mathcal{X}_1, \ldots, \mathcal{X}_n$ the conditional expectation value for the next observation $\mathrm{E}(\mathcal{X}_{n+1})$ is equal to the last recorded value $\mathcal{X}_n$.

A continuous time martingale refers to a random variable $\mathcal{X}(t)$ with the expectation value $\mathrm{E}\big(\mathcal{X}(t)\big)$. We define first the conditional expectation value of the random variable for $\mathcal{X}(t_0) = x_0$ and $\mathrm{E}(|\mathcal{X}(t)|) < \infty$:

$$\mathrm{E}\big(\mathcal{X}(t)|(x_0, t_0)\big) \;\doteq\; \int \mathrm{d}x\, p\,(x, t|x_0, t_0) \;.$$

In a martingale the conditional mean is simply given by

$$\mathrm{E}\big(\mathcal{X}(t)|(x_0, t_0)\big) \;= x_0 \;. \tag{3.8}$$

The mean value at time t is identical to the initial value of the process. The martingale property is rather strong and we shall use it for several specific situations.

As an example of a martingale we show the unlimited symmetric random walk in one dimension (figure 3.3): Equally sized steps of length l to the right and to the left are taken with equal probability. In the discrete time random walk the waiting time between two steps is θ and appropriately we measure time in multiples of the waiting time: $t - t_0 = k \cdot \theta$, and the position in multiples of the step length l. The corresponding probability to be at location $x - x_0 = n \cdot l$ at time θ is simply expressed in doubles of variables (n, k)

$$\begin{aligned} P(n, k+1\,|\,n_0, k_0) \;&=\; \frac{1}{2}\Big(P(n+1, k\,|\,n_0, k_0) + P(n-1, k\,|\,n_0, k_0)\Big) \;, \\ P_{n,k+1} \;&=\; \frac{1}{2}\big(P_{n+1,k} + P_{n-1,k}\big) \quad \text{with} \quad P_{n,k_0} = \delta_{n,n_0} \;, \end{aligned} \tag{3.9}$$

[9] For convenience we change here the numbering of times and apply the notation of equation (3.3')

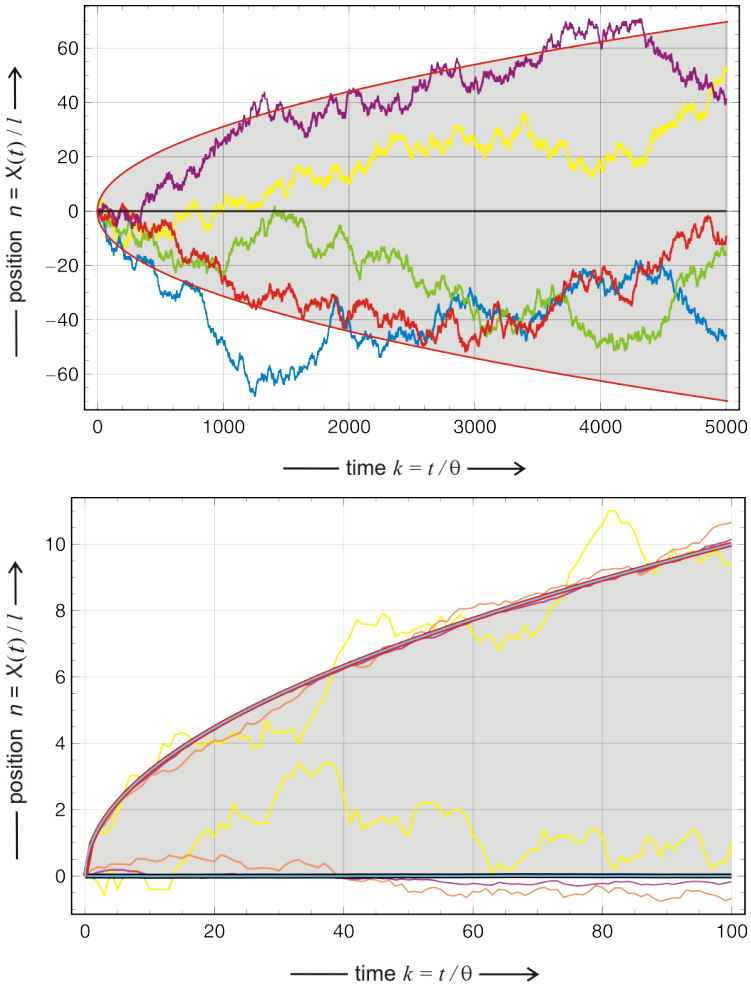


Fig. 3.3 The discrete time one-dimensional random walk. The random walk in one dimension on an infinite line, $x \in \mathbb{R}$, is shown as an example of a martingale. The upper part shows five trajectories, $\mathcal{X}(t)$, which were calculated with different seeds for the random number generator. The expectation value $\mathrm{E}\big(\mathcal{X}(t)\big) = x_0 = 0$ is constant (black line), the variance grows linearly with time $\mathrm{var}\big(\mathcal{X}(t)\big) = k = t/\theta$, and the standard deviation is $\sigma\big(\mathcal{X}(t)\big) = \sqrt{k}$. The two red lines correspond the one standard deviation band $\mathrm{E}(t) \pm \sigma(t)$, and the gray area represent the confidence interval of 68,2 %. Choice of parameters: $\theta^{-1} = 1\,[t.u.]$ $(= 2\vartheta)$; $l = 1\,[l.u.]$; random number generator: *Mersenne Twister*, seeds: 491 (yellow), 919 (blue), 023 (green), 877 (red), 127 (violet). The lower part of the figure shows the convergence of sample mean and sample standard deviation according to equation (3.4) with increasing number N of sampled trajectories: $N = 10$ (yellow), 100 (orange), 1000 (purple), and 10^6 (red and black). The last curve is almost indistinguishable from the limit $N \to \infty$ (ice blue line upon the red and the black curve). Parameters are the same as in the upper part, *Mersenne Twister*, seeds: 637

where the *short-hand* notation expresses the initial conditions in a separate equation. Our choice of variables allows for simplified initial conditions without loosing generality: $n_0 = 0$ and $k_0 = 0$. Equation (3.9) can be readily solved by means of the characteristic function

$$\phi(s,k) \;=\; \mathrm{E}\big(e^{\dot{\imath}ns}\big) \;=\; \sum_{n=-\infty}^{\infty} P(n,k\,|\,0,0,)\,e^{\dot{\imath}ns} = \sum_{n=-\infty}^{\infty} P_{n,k}\,e^{\dot{\imath}ns}\;. \quad (2.29')$$

Implementation of equation (3.9) yields

$$\phi(s,k+1) \;=\; \frac{1}{2}\phi(s,k)\big(e^{\dot{\imath}s} + e^{-\dot{\imath}s}\big) \;=\; \cosh(\dot{\imath}s) \;\;\text{with}\;\; \phi(s,0) = 1\;,$$

and the solution is calculated to be

$$\begin{aligned}\phi(s,k) \;&=\; \cosh^k(\dot{\imath}s) \;= \\ &=\; \frac{1}{2^k}\left(e^{\dot{\imath}ks} + \binom{k}{1}e^{\dot{\imath}(k-2)s} + \binom{k}{2}e^{\dot{\imath}(k-4)s} + \ldots + e^{-\dot{\imath}ks}\right). \end{aligned} \quad (3.10a)$$

Equating the coefficients for the individual $e^{\dot{\imath}ns}$ terms in expressions (3.10a) and (2.29') determines the probabilities

$$P_{n,k} \;=\; \begin{cases} \frac{1}{2^k}\binom{k}{\nu} & \text{if } |n| \le k, \nu = \frac{k-n}{2} \in \mathbb{N}, \\ 0 & \text{otherwise}\;. \end{cases} \quad (3.10b)$$

The distribution is a binomial distribution with $k+1$ terms, a width $2k$ and every second term being equal to zero. It is spreading with time according to $t = k \cdot \theta$.

Calculation of first and second moments is straightforward and can be achieved best by using the derivatives of the characteristic function as shown in equation (2.31):

$$\begin{aligned}\frac{\partial\phi(s,k)}{\partial s} \;&=\; \dot{\imath}\,n\,\cosh^{n-1}(\dot{\imath}s)\cdot\sinh(\dot{\imath}s) \;\;\text{and} \\ \frac{\partial^2\phi(s,k)}{\partial s^2} \;&=\; -n\left(\cosh^n(\dot{\imath}s) + (n-1)\,\cosh^{n-2}(\dot{\imath}s)\cdot\sinh^2(\dot{\imath}s)\right)\end{aligned}$$

Insertion of $s = 0$ yields $(\partial\phi/\partial s)|_{s=0} = 0$ and $(\partial^2\phi/\partial s^2)|_{s=0} = -n$ and by equation (2.31) we obtain with $n(0) = n_0$ and $k(0) = k_0$ for the moments

$$\mathrm{E}\big(\mathcal{X}(t)\big) \;=\; x_0 \;=\; n_0 \cdot l \quad\text{and}\quad \mathrm{var}\big(\mathcal{X}(t)\big) \;=\; t - t_0 \;=\; (k-k_0)\,\theta\;. \quad (3.11)$$

The unlimited, symmetric and discrete random walk in one dimension is a martingale and the standard deviation $\sigma\big(\mathcal{X}(t)\big)$ increases with $\sqrt{t}$ as predicted in the path-breaking works of Albert Einstein [108] and Marian von

Smoluchowski [453]. This implies that trajectories in general will diverge and approach $\pm\infty$.

We remark that the standardized sum of the outcomes of Bernoulli trials, $s(n) = \mathcal{S}(n)/n - \mu$ with $\mathcal{S}_n = \sum_{i=1}^{n} \mathcal{X}_i$ and $\mathcal{X}_i = \pm 1$, which had been used to illustrate the law of the iterated logarithm (figure 2.12), represents a martingale as well but here the trajectories are confined by the domain $s(n) = \pm 1$ and the long-term limit is zero. A time scale in this case results from the assignment of a time interval between two successive trials.

The somewhat relaxed notion of a *semimartingale* is of importance because it covers the majority of processes that are accessible to modeling by *stochastic differential equations*. A semimartingale is composed of a *local martingale* and an *adapted càdlàg-process*[10] with bounded variation

$$\mathcal{X}(t) \;=\; \mathcal{M}(t) \,+\, \mathcal{A}(t)$$

A local martingale is a stochastic process that satisfies locally the martingale property (3.8) but its expectation value $\langle \mathcal{M}(t) \rangle$ may be distorted at long times by large values of low probability. Hence, every martingale is a local martingale and every bounded local martingale is a martingale. In particular, every driftless diffusion process is a local martingale but need not be a martingale.

An adapted process $\mathcal{A}(t)$ is *nonanticipating* in the sense that it *cannot see into the future*. An informal interpretation [464, section II.25] would say: A stochastic process $\mathcal{X}(t)$ is adapted if and only if for every realization and for every time t, $\mathcal{X}(t)$ is known at time t and not before. The notion 'nonanticipating' is irrelevant for deterministic processes but matters for processes containing fluctuating elements, because the independence of random or irregular increments makes it impossible to look into the future. The concept of adapted processes is essential for the definition and evaluation of the Itō stochastic integral, which is based on the assumption that the integrand is an adapted process (section 3.4.2).

Two generalizations of martingales are in common use: (i) A discrete time *submartingale* is a sequence $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, \ldots$, of random variables that satisfy

$$\mathrm{E}(\mathcal{X}_{n+1} | \mathcal{X}_1, \ldots, \mathcal{X}_n) \;\geq\; \mathcal{X}_n \;, \tag{3.12}$$

and for the continuous time analogue we have the condition

$$\mathrm{E}\big(\mathcal{X}(t) | \{\mathcal{X}(\theta) : \theta \leq s\}\big) \;\geq \mathcal{X}(s) \;\forall\; s \,\leq\, t \;. \tag{3.13}$$

(ii) The relations for *supermartingales* are in complete analogy to the submartingales, only the '$\geq$' relations have to be replaced by '$\leq$':

[10] The property *càdlàg* is an acronym from French for "*continue à droite, limites à gauche*". It is a common property of step functions in probability theory (section 1.6.2).

$$\mathrm{E}(\mathcal{X}_{n+1}|\mathcal{X}_1, \ldots, \mathcal{X}_n) \;\leq\; \mathcal{X}_n \;, \tag{3.14}$$

$$\mathrm{E}\big(\mathcal{X}(t)|\{\mathcal{X}(\theta) : \theta \leq s\}\big) \;\leq \mathcal{X}(s) \;\forall\; s \leq t \;. \tag{3.15}$$

A straightforward consequence of the property of martingales is: If a sequence or a function of random variables is a simultaneously submartingale and a supermartingale it is a martingale.

3.1.3.3 Markov processes

Markov processes are processes that share the Markov property, which in a nutshell assumes that knowledge of the present alone is all we need to predict the future of such a process, or in other words information on the past will not improve the prediction the future. Although processes that fulfil the Markov property are only a minority among stochastic processes[440] in general, they are of particular importance because most models in science assume the Markov property, and this assumption facilitates the analysis enormously.

The Markov process is named after the Russian mathematician Andrey Markov[11] and can formulated straightforwardly in terms of conditional probabilities:

$$p\,(x_1, t_1; x_2, t_2; \cdots | x_k, t_k; x_{k+1}, t_{k+1}, \cdots) \;=\; p\,(x_1, t_1; x_2, t_2; \cdots | x_k, t_k) \;. \tag{3.16}$$

As said the Markov condition expresses independence of the history of the process prior to time t_k. For example, we have

$$p\,(x_1, t_1; x_2, t_2; x_3, t_3) \;=\; p\,(x_1, t_1 | x_2, t_2)\, p\,(x_2, t_2 | x_3, t_3) \;.$$

As we have seen in section 1.6.4 any arbitrary joint probability can be simply expressed as products of conditional probabilities:

$$\begin{aligned} &p\,(x_1, t_1; x_2, t_2; x_3, t_3; \cdots; x_n, t_n) \;= \\ &= \; p\,(x_1, t_1 | x_2, t_2)\, p\,(x_2, t_2 | x_3, t_3) \cdots p\,(x_{n-1}, t_{n-1} | x_n, t_n)\, p\,(x_n, t_n) \end{aligned} \tag{3.16'}$$

under the assumption of time ordering $t_1 \geq t_2 \geq t_3 \geq \ldots \geq t_{n-1} \geq t_n$. Because these products of conditional probabilities of two events one speaks also of a *Markov chain*.

The Bernoulli process (section 3.1.3.1) can be seen now as a special Markov process, in which the next state is not only independent of the past states but also of the current state.

---

[11] The Russian mathematician Andrey Markov (1856-1922) is one of the founders of Russian probability theory and pioneered the concept of memory free processes, which is named after him. He expressed more precisely the assumptions that were made by Albert Einstein [108] and Marian von Smoluchowski [453] in their derivation of the diffusion process.

3.1.3.4 Stationarity

Stationarity for a deterministic process implies that all observable time dependencies become zero at stationary states. In the case of multistep processes the definition leaves two possibilities open: (i) At thermodynamic equilibrium the fluxes of all individual steps vanish as it is expressed in the *principle of detailed balance* [430], or (ii) only the total flux that is the the sum of all fluxes becomes zero. Stationarity of stochastic processes in general, and of Markov processes in particular, is more subtle since random fluctuations do not vanish at equilibrium. Several definitions of stationarity are possible, three of them are relevant for our purposes here.

Strong stationarity. A stochastic process is called strictly or *strongly stationary* if $\mathcal{X}(t)$ and $\mathcal{X}(t+\Delta t)$ obey the same statistics for every Δt. Accordingly, joint probability densities are invariant to translation in time:

$$p\,(x_1,t_1;x_2,t_2;\ldots;x_n,t_n) \;=\; p\,(x_1,t_1+\Delta t;x_2,t_2+\Delta t;\ldots;x_n,t_n+\Delta t)\;. \quad (3.17)$$

In other words, the probabilities are only functions of time differences, $\Delta t = t_k - t_j$, and this leads to time independent stationary *one-time probabilities*

$$p\,(x,t) \implies \bar{p}\,(x)\;, \quad (3.18)$$

and two-times joint or conditional probabilities of the form

$$\begin{aligned} p\,(x_1,t_1;x_2,t_2) &\implies \bar{p}\,(x_1,t_1-t_2;x_2,0) \;\text{ and} \\ p\,(x_1,t_1|\,x_2,t_2) &\implies \bar{p}\,(x_1,t_1-t_2|\,x_2,0)\;. \end{aligned} \quad (3.19)$$

Since all joint probabilities of a Markov process can be written as products of two-time conditional probabilities and a one-time probability (3.16'), the necessary and sufficient condition for stationarity is cast into the requirement to be able to write all one- and two-time probabilities as shown in equations (3.18) and (3.19). A Markov process that becomes stationary in the limit $t \to \infty$ or $t_0 \to -\infty$ is called a *homogeneous Markov process*.

Weak stationarity. The notion of *weak stationarity* or covariance stationarity is used in signal processing and relaxes the stationarity condition (3.17) for a process $\mathcal{X}(t)$ to

$$\begin{aligned} \mathrm{E}\Big(\mathcal{X}(t)\Big) \;&=\; \mu_{\mathcal{X}}(t) \;=\; \mu_{\mathcal{X}}(t+\Delta t)\;\forall\;\Delta t \in \mathbb{R} \;\text{ and} \\ \mathrm{cov}\Big(\mathcal{X}(t_1),\mathcal{X}(t_2)\Big) \;&=\; \mathrm{E}\Big(\big(\mathcal{X}(t_1)-\mu_{\mathcal{X}}(t_1)\big)\big(\mathcal{X}(t_2)-\mu_{\mathcal{X}}(t_2)\big)\Big) \;= \\ &=\; \mathrm{E}\Big(\mathcal{X}(t_1)\mathcal{X}(t_2)\Big) \;-\; \mu_{\mathcal{X}}(t_1)\mu_{\mathcal{X}}(t_2) \;= \\ &=\; C_{\mathcal{X}}(t_1,t_2) \;=\; C_{\mathcal{X}}(t_1-t_2,0) \;=\; C_{\mathcal{X}}(\Delta t)\;. \end{aligned} \quad (3.20)$$

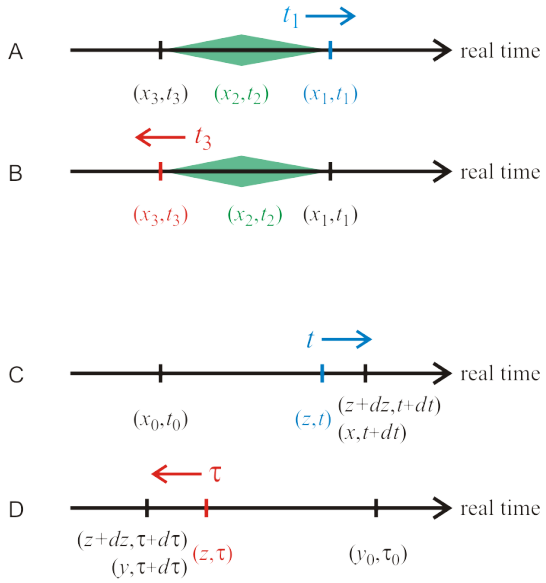


Fig. 3.4 Notation of time dependent variables. In the following sections we shall require several time dependent variables and adopt the following notations: In case of the Chapman-Kolmogorov equation we require three variables at different times denoted by x_1, x_2, and x_3. The variable x_2 is associated with intermediate time t_2 (green) and disappears through integration. In the forward equation (x_3, t_3) are fixed initial conditions and (x_1, t_1) is moving (A). For backward integration the opposite relation is assumed: (x_1, t_1) being fixed and (x_3, t_3) moving (B). In both cases *real time* progresses from the left to the right, computational time increases in the same direction as real time in the forward evaluation (blue) but in opposite direction for backward evaluation (red). The lower part of the figure shows notations used for forward and backward differential Chapman-Kolmogorov equations: In the forward equation (C) $x(t)$ is the variable, the initial conditions are denoted by (x_0, t_0) and (z, t) is an intermediate double. In the backward equation the time order is reversed (D): $y(\tau)$ is the variable and (y_0, τ_0) are the final conditions. In both cases we could use $z + dz$ instead x or y, respectively, but the equations would be less clear then.

Instead of the entire probability function only the mean of the process, $\mu_{\mathcal{X}}$, must be constant and the *autocovariance function*[12] of the stochastic process $\mathcal{X}(t)$ denoted by $C_{\mathcal{X}}(t_1, t_2)$ does not depend on t_1 and t_2 but only on the difference $\Delta t = t_1 - t_2$.

Second order stationarity. The notion of *second order stationarity* of a process with finite mean and finite autocovariance expresses the fact that the conditions of strict stationarity are applied only to pairs of random variables

[12] The notion *autocovariance* reflects the fact the process is correlated with itself at another time and cross-covariance implies the correlation of two different processes (for the relation between autocorrelation and autocovariance see section 3.1.4).

from the time series and then the first and second order density functions fulfil:

$$\begin{aligned} f_{\mathcal{X}}(x_1;t_1) &= f_{\mathcal{X}}(x_1;t_1+\Delta t)\ \forall\ (t_1,\Delta t) \ \text{ and} \\ f_{\mathcal{X}}(x_1,x_2;t_1,t_2) &= f_{\mathcal{X}}(x_1,x_2;t_1+\Delta t,t_2+\Delta t)\ \forall\ (t_1,t_2,\Delta t)\ . \end{aligned} \tag{3.21}$$

The definition can be extended to higher orders and then strict stationarity is tantamount to stationarity in all orders. A second order stationary process fulfils the criteria for weak stability but a process can be stationary in the wide sense without fulfilling the criteria of second order stationarity.

3.1.3.5 Continuity in stochastic processes

Continuity in deterministic processes requires absence of any kind of jumps but does not require differentiability expressed as continuity in the first derivative. We recall the conventional definition of continuity at $x = x_0$:

$$\forall\,\varepsilon > 0\ \exists\ \delta > 0 \text{ such that } \forall\, x:\ |x - x_0| < \delta \implies |f(x) - f(x_0)| < \varepsilon\,.$$

In other words there is the requirement that $|f(x) - f(x_0)|$ can become arbitrarily small for all $|x - x_0|$ no matter how close x is to x_0 – no jumps are allowed. The condition of continuity in Markov processes is defined analogously but requires a more detailed discussion. For this goal we consider a process that progresses from location z at time t to location $x = z + \Delta z$ at time $t+\Delta t$ denoted as $(z,t) \to (z+\Delta z, t+\Delta t) = (x, t+\Delta t)$.[13] The process is continuous if and only if in the limit $\Delta t \to 0$ the probability of z to be finitely different from x goes to zero faster than Δt as expressed by the equation

$$\lim_{\Delta t\to 0}\ \frac{1}{\Delta t}\int_{|\Delta z|=|x-z|>\varepsilon} \mathrm{d}x\, p\,(x, t+\Delta t|z,t)\ =\ 0\,, \tag{3.22}$$

and this convergence is uniform in z, t, and Δt. In other words, the difference in probability as a function of $|z - x\,|$ approaches zero sufficiently fast and therefore no jumps occur in the random variable $\mathcal{X}(t)$.

For the analysis of continuity in Markov processes we choose two illustrative examples [157, pp. 65-68], which have the trajectories sketched in figure 3.5: (i) the Brownian motion [55] or Wiener process, which is the continuous version of the random walk in one dimension shown in figure 3.3[14] and which leads to a normally distributed conditional probability,

$$p\,(x, t+\Delta t|z,t)\ =\ \frac{1}{\sqrt{4\pi D\Delta t}}\ \exp\left(-\frac{(x-z)^2}{4D\Delta t}\right)\ , \tag{3.23}$$

[13] The notation used for time dependent variables is listed in figure 3.4. For convenience and readability we write x for $z + \Delta z$.

[14] Later on we shall discuss the limit of the random walk for vanishing step size in more detail and call it a *Wiener process* (section3.2.2.2).

and (ii) the so-called Cauchy process following the Cauchy-Lorentz distribution,

$$p\,(x, t+\Delta t|z,t) \;=\; \frac{1}{\pi}\,\frac{\Delta t}{(x-z)^2+\Delta t^2}\;. \tag{3.24}$$

The distribution in case of the Wiener process follows directly from the binomial distribution of the random walk (3.10b) in the limit of vanishing step size. For the analysis of continuity we exchange the limit and the integral, introduce $\vartheta = (\Delta t)^{-1}$, perform the limit $\vartheta \to \infty$, and find

$$\begin{aligned}
&\lim_{\Delta t\to 0}\,\frac{1}{\Delta t}\int_{|x-z|>\varepsilon}\mathrm{d}x\,\frac{1}{\sqrt{4\pi D}}\,\frac{1}{\sqrt{\Delta t}}\,\exp\left(-\frac{(x-z)^2}{4D\Delta t}\right) \;= \\
&= \int_{|x-z|>\varepsilon}\mathrm{d}x\,\lim_{\Delta t\to 0}\,\frac{1}{\Delta t}\,\frac{1}{\sqrt{4\pi D}}\,\frac{1}{\sqrt{\Delta t}}\,\exp\left(-\frac{(x-z)^2}{4D\Delta t}\right) \;= \\
&= \int_{|x-z|>\varepsilon}\mathrm{d}x\,\lim_{\vartheta\to\infty}\,\frac{1}{\sqrt{4\pi D}}\,\frac{\vartheta^{3/2}}{\exp\left(\frac{(x-z)^2}{4D}\,\vartheta\right)}\,, \quad \text{where} \\
&\lim_{\vartheta\to\infty}\,\frac{\vartheta^{3/2}}{1+\frac{(x-z)^2}{4D}\cdot\vartheta+\frac{1}{2!}\left(\frac{(x-z)^2}{4D}\right)^2\cdot\vartheta^2+\frac{1}{3!}\left(\frac{(x-z)^2}{4D}\right)^3\cdot\vartheta^3+\ldots} \;=\; 0\;.
\end{aligned}$$

Since the power expansion of the exponential in the denominator increases faster than every finite power of ϑ, the ratio vanishes in the limit $\vartheta \to \infty$, the value of the integral is zero, and the Wiener process is continuous everywhere. Although it is continuous, the trajectory of the Wiener Process is extremely irregular since it is nowhere differentiable (figure 3.5).

In the second example, the Cauchy process, we exchange limit and integral as in case of the Wiener process, and perform the limit $\Delta t \to 0$:

$$\begin{aligned}
&\lim_{\Delta t\to 0}\,\frac{1}{\Delta t}\int_{|x-z|>\varepsilon}\mathrm{d}x\,\frac{\Delta t}{\pi}\,\frac{1}{(x-z)^2+\Delta t^2} \;= \\
&= \int_{|x-z|>\varepsilon}\mathrm{d}x\,\lim_{\Delta t\to 0}\,\frac{1}{\Delta t}\,\frac{\Delta t}{\pi}\,\frac{1}{(x-z)^2+\Delta t^2} \;= \\
&= \int_{|x-z|>\varepsilon}\mathrm{d}x\,\lim_{\Delta t\to 0}\,\frac{1}{\pi}\,\frac{1}{(x-z)^2+\Delta t^2} \;=\; \int_{|x-z|>\varepsilon}\frac{1}{\pi(x-z)^2}\,\mathrm{d}x \;\neq\; 0\;.
\end{aligned}$$

The value of the last integral, $I = \int_{|x-z|>\varepsilon}^{\infty} dx/(x-z)^2 = 1/\big(\pi(x-z)\big)$, is of the order $I \approx 1/\varepsilon$ and accordingly finite. Consequently, the curve for the Cauchy-process is not only irregular but also discontinuous (figure 3.5).

Both processes, as required for consistency, fulfill the relation

$$\lim_{\Delta t\to 0}\,p\,(x, t+\Delta t|\,z,t) \;=\; \delta(x-z)\;, \tag{3.25}$$

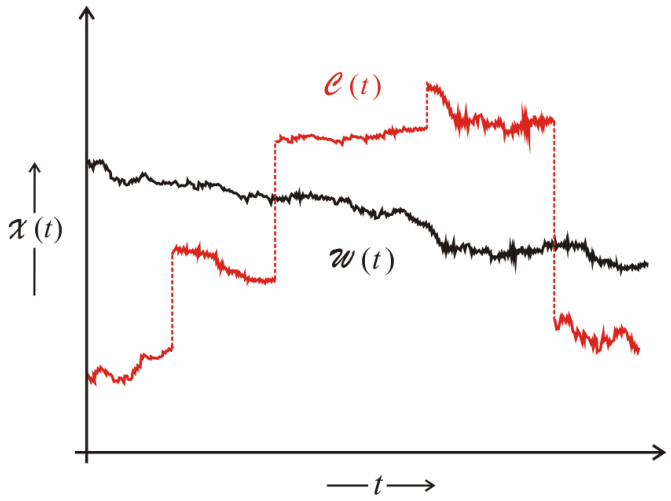


Fig. 3.5 Continuity in Markov processes. Continuity is illustrated by means of two stochastic processes of the random variable $\mathcal{X}(t)$, the Wiener process $\mathcal{W}(t)$ (3.23; black) and the Cauchy process $\mathcal{C}(t)$ (3.24; gray). The Wiener process describes Brownian motion and is continuous but almost nowhere differentiable. The even more irregular Cauchy process is wildly discontinuous.

where $\delta(\cdot)$ is the so-called delta-function (see section 1.6.3). We are now in the position to give a concise mathematical definition for continuity in Markov processes [157, p. 46], which will be used to derive a comprehensive and convenient equation for stochastic processes. For general validity we use vector notation for the locations:

A Markov process has – with probability one – sample paths that are continuous functions of time t, if for any $\varepsilon > 0$ the limit

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} \mathrm{d}\boldsymbol{x}\, p\left(\boldsymbol{x}, t+\Delta t | \, \boldsymbol{z}, t\right) \;=\; 0 \;. \tag{3.26}$$

is approached uniformly in $\boldsymbol{z}$, t, and Δt.

In essence, equation (3.26) expresses the fact that probabilistically the difference between $\boldsymbol{x}$ and $\boldsymbol{z}$ converges to zero faster than Δt does.

3.1.4 Autocorrelation functions and spectra

Analysis of experimentally recorded or computer created trajectories is often largely facilitated by the usage of additional tools complementing moments and probability distributions since they can, in principle, be derived from single recordings. These tools are autocorrelation functions and spectra of random variables, which provide direct insight into the dynamics of the process, since they are dealing with relations between points collected at different times from the same sample path. The autocorrelation is readily accessible experimentally (see e.g. the application of the autocorrelation function to fluorescence correlation spectroscopy see section 4.4.2) and represents a basic tool in *time series analysis* (see, for example, [458]).

Convolution, cross-correlation and autocorrelation. These three integral relations between functions $f(t)$ and $g(t)$ are important in statistics, in particular in signal processing. The convolution is defined as

$$(f * g)(\mathbf{x}) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \mathrm{d}\mathbf{y}\, f(\mathbf{y})\, g(\mathbf{x} - y) \;=\; \int_{-\infty}^{\infty} \mathrm{d}\mathbf{y}\, f(\mathbf{x} - \mathbf{y})\, g(\mathbf{y}) \;, \quad (3.27)$$

where $\mathbf{x}$ and $\mathbf{y}$ are understood as vectors in n-dimensional space, $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n$. Among other important properties the *convolution theorem* is of great practical importance because it allows for straightforward computation of the convolution as the product of two integrals after Fourier transform:

$$\mathcal{F}(f * g) \;=\; \mathcal{F}(f) \cdot \mathcal{F}(g) \;\text{ and }\; f * g \;=\; \mathcal{F}^{-1}\big(\mathcal{F}(f) \cdot \mathcal{F}(g)\big) \;, \quad (3.28)$$

where the Fourier transform and its inverse are defined by[15]

$$\tilde{f}(\boldsymbol{\xi}) \;=\; \mathcal{F}(f) \;=\; \int_{-\infty}^{\infty} f(\mathbf{x})\, \exp\left(-2\pi i\, \mathbf{x} \cdot \boldsymbol{\xi}\right) \mathrm{d}\mathbf{x} \;,$$

$$f(\mathbf{x}) \;=\; \mathcal{F}^{-1}(\tilde{f}) \;=\; \int_{-\infty}^{\infty} \tilde{f}(\boldsymbol{\xi})\, \exp\left(2\pi i\, \mathbf{x} \cdot \boldsymbol{\xi}\right) \mathrm{d}\boldsymbol{\xi} \;.$$

The convolution theorem can also be inverted and yields

$$\mathcal{F}(f \cdot g) \;=\; \mathcal{F}(f) * \mathcal{F}(g) \;.$$

The *cross correlation* is related to the convolution and commonly defined as

$$(f \star g)(\mathbf{x}) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \mathrm{d}\mathbf{y}\, f^*(\mathbf{y})\, g(\mathbf{x} + \mathbf{y}) \;, \quad (3.29)$$

and in analogy to the convolution theorem the relation

[15] We remark that this definition of the Fourier transform differs from the conventional one in modern physics (see section 2.2.3).

$$\mathcal{F}(f \star g) \;=\; \big(\mathcal{F}(f)\big)^* \cdot \mathcal{F}(g)$$

holds for the Fourier transform of the cross-correlation. The autocorrelation,

$$(f \star f)(\mathbf{x}) \;\stackrel{\text{def}}{=}\; \int_{-\infty}^{\infty} \mathrm{d}\mathbf{y}\, f^*(\mathbf{y})\, f(\mathbf{x}+\mathbf{y}) \;, \tag{3.30}$$

is a special case of cross-correlation, the cross-correlation of a function f with itself.

Autocorrelation and spectrum. The autocorrelation function of a stochastic process is defined by (2.9') as the coefficient of correlation $\rho(\mathcal{X},\mathcal{Y})$ of the random variable at some time, $\mathcal{X} = \mathcal{X}(t_1)$, with the variable at another time, $\mathcal{Y} = \mathcal{X}(t_2)$:

$$\begin{aligned} R(t_1,t_2) \;&=\; \rho\Big(\mathcal{X}(t_1),\mathcal{X}(t_2)\Big) \;= \\ &=\; \frac{\mathrm{E}\Big(\big(\mathcal{X}(t_1)-\mu_{\mathcal{X}}(t_1)\big)\big(\mathcal{X}(t_2)-\mu_{\mathcal{X}}(t_2)\big)\Big)}{\sigma_{\mathcal{X}}(t_1)\,\sigma_{\mathcal{X}}(t_2)} \;,\; R \in [-1,1] \;. \end{aligned} \tag{3.31}$$

Thus the autocorrelation function is obtained from the autocovariance function (3.20) through division by the product of the standard deviations:

$$R(t_1,t_2) \;=\; \frac{\mathrm{cov}\big(\mathcal{X}(t_1)\mathcal{X}(t_2)\big)}{\sigma_{\mathcal{X}}(t_1)\,\sigma_{\mathcal{X}}(t_2)} \;.$$

Accordingly, the autocorrelation of the random variable $\mathcal{X}(t)$ is a measure of the influence the value of $\mathcal{X}$ recorded time t_1 has on the measurement of the same variable at time t_2. Under the assumption that we are dealing with a weak or second order stationary process the mean and the variance are independent of time and then the autocorrelation function depends only on the time difference $\Delta t = t_2 - t_1$:

$$R(\Delta t) \;=\; \frac{\mathrm{E}\Big(\big(\mathcal{X}(t)-\mu_{\mathcal{X}}\big)\big(\mathcal{X}(t+\Delta t)-\mu_{\mathcal{X}}\big)\Big)}{\sigma_{\mathcal{X}}^2} \;,\; R \in [-1,1] \;. \tag{3.31'}$$

In spectroscopy the autocorrelation obtained for a spectroscopic signal $\mathcal{F}(t)$ measured as a function of time. Then we are dealing with

$$\begin{aligned} G(\Delta t) \;&=\; \langle \mathcal{F}(t)\mathcal{F}(t+\Delta t)\rangle \;=\; \mathrm{E}\Big(\mathcal{F}(t)\mathcal{F}(t+\Delta t)\Big) \;= \\ &=\; \lim_{t\to\infty} \frac{1}{t} \int_0^t \mathrm{d}\theta\, x(\theta)\, x(\theta+\Delta t) \;. \end{aligned} \tag{3.32}$$

Thus, the autocorrelation function is the time average of the product of two values recorded at different times with a given interval Δt.

Another relevant quantity is the *spectrum* or the spectral density of the quantity $x(t)$. In order to derive the spectrum, we construct a new variable $y(\omega)$ by means of the transformation $y(\omega) = \int_0^t \mathrm{d}\theta\, e^{\mathrm{i}\omega\theta}\, x(\theta)$. The spectrum is then obtained from y by performing the limit $t \to \infty$:

$$S(\omega) \;=\; \lim_{t\to\infty} \frac{1}{2\pi t}\, |y(\omega)|^2 \;=\; \lim_{t\to\infty} \frac{1}{2\pi t} \left| \int_0^t \mathrm{d}\theta\, e^{\mathrm{i}\omega\theta}\, x(\theta) \right|^2 . \qquad (3.33)$$

The autocorrelation function and the spectrum are closely related. By some calculations one finds

$$S(\omega) \;=\; \lim_{t\to\infty} \left(\frac{1}{\pi} \int_0^t \cos(\omega\tau)\, \mathrm{d}\tau\; \frac{1}{t} \int_0^{t-\tau} x(\theta)\, x(\theta+\tau)\, \mathrm{d}\theta \right) .$$

Under certain assumptions, which ensure the validity of the interchanges of order, we may take the limit $t \to \infty$ and find

$$S(\omega) \;=\; \frac{1}{\pi} \int_0^\infty \cos(\omega\tau)\, G(\tau)\, \mathrm{d}\tau \;.$$

This result relates the Fourier transform of the autocorrelation function to the spectrum and can be cast in an even prettier form by using

$$G(-\tau) \;=\; \lim_{t\to\infty} \frac{1}{t} \int_{-\tau}^{t-\tau} \mathrm{d}\theta\, x(\theta)\, x(\theta+\tau) \;=\; G(\tau)$$

to yield the *Wiener-Khinchin theorem* named after the American physicist Norbert Wiener and the Russian mathematician Aleksandr Khinchin

$$S(\omega) \;=\; \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\mathrm{i}\omega\tau}\, G(\tau)\, \mathrm{d}\tau \;\text{ and }\; G(\tau) \;=\; \int_{-\infty}^{+\infty} e^{\mathrm{i}\omega\tau} S(\omega)\, \mathrm{d}\omega \;. \qquad (3.34)$$

Spectrum and autocorrelation function are related to each other by the Fourier transformation and its inversion.

Equation (3.34) allows for a straightforward proof that the Wiener process $\vec{\mathcal{W}}(t) = W(t)$ gives rise to *white noise* (subsection 3.2.2.2). Let $\mathbf{w}$ be a zero-mean random vector with the identity matrix as (auto)covariance or autocorrelation matrix:

$$\mathrm{E}(\mathbf{w}) \;=\; \boldsymbol{\mu} \;=\; \mathbf{0} \;\text{ and }\; \mathrm{cov}(\mathcal{W},\mathcal{W}) \;=\; \mathrm{E}(\mathbf{w}\mathbf{w}') \;=\; \mathrm{var}\,\mathbb{I} \;,$$

then the Wiener process $W(t)$ fulfils the relations,

$$\begin{aligned} \mu_W(t) \;&=\; \mathrm{E}\big(\mathcal{W}(t)\big) \;=\; 0 \;\text{ and} \\ G_W(\tau) \;&=\; \mathrm{E}\big(\mathcal{W}(t)\mathcal{W}(t+\tau)\big) \;=\; \delta(\tau) \;, \end{aligned}$$

defining it as a zero-mean process with infinite power at zero time shift. For the spectral density of the Wiener process we obtain:

$$S_W(\omega) \;=\; \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\mathrm{i}\omega\tau}\,\delta(\tau)\,\mathrm{d}\tau \;=\; \frac{1}{2\pi}\;. \tag{3.35}$$

The spectral density of the Wiener process is a constant and hence all frequencies in the noise are represented with equal weight. Mixing all frequencies in electromagnetic radiation with equal weight yields white light and this property of visible light gave the name for *white noise*. In *colored noise* the noise frequencies do not meet the condition of the uniform distribution. Pink or flicker noise, for example, has a spectrum close to $S(\omega) \propto \omega^{-1}$ and red or Brownian noise fulfils $S(\omega) \propto \omega^{-2}$.

The time average of a signal as expressed by an autocorrelation function is complemented by the *ensemble average*, $\langle\cdot\rangle$, or the expectation value of the corresponding random variable, $\mathrm{E}(\cdot)$, which implies an (infinite) number of repeats of the same measurement. If the assumption of *ergodic behavior* is true, the time average is equal to the ensemble average. Thus we find for a fluctuating quantity $\mathcal{X}(t)$ in the ergodic limit

$$\mathrm{E}\big(\mathcal{X}(t),\mathcal{X}(t+\tau)\big) \;=\; \langle x(t)x(t+\tau)\rangle \;=\; G(\tau)\;.$$

It is straightforward to consider dual quantities which are related by Fourier transformation and get:

$$x(t) \;=\; \frac{1}{2\pi}\int \mathrm{d}\omega\, c(\omega)\, e^{\mathrm{i}\omega t} \quad\text{and}\quad c(\omega) \;=\; \int \mathrm{d}t\, x(t)\, e^{-\mathrm{i}\omega t}\;.$$

We use this relation to derive several important results. Measurements refer to real quantities $x(t)$ and this implies: $c(\omega) = c^*(-\omega)$. From the condition of stationarity follows $\langle x(t)x(t')\rangle = f(t-t')$ and hence it depends on τ only and does not depend on t, and we derive

$$\begin{aligned}\langle c(\omega)c^*(\omega')\rangle \;&=\; \frac{1}{(2\pi)^2}\iint \mathrm{d}t\,\mathrm{d}t'\, e^{-\mathrm{i}\omega t+\mathrm{i}\omega t'}\;\langle x(t)x(t')\rangle \;=\\ &=\; \frac{\delta(\omega-\omega')}{2\pi}\int \mathrm{d}\tau\, e^{\mathrm{i}\omega\tau}\, G(\tau) \;=\; \delta(\omega-\omega')\, S(\omega)\;.\end{aligned}$$

The last expression relates not only the mean square $\big\langle |c(\omega)|^2\big\rangle$ with the spectrum of the random variable, it shows also that stationarity alone implies that $c(\omega)$ and $c^*(\omega')$ are uncorrelated.

3.2 The Chapman-Kolmogorov equation

At the basis of general modeling of stochastic processes stands a straightforward consideration concerning the propagation of probability distributions in time, the question how to calculate the probability to come from the random variable $\mathcal{N}_3 = n_3$ at time $t = t_3$ to $\mathcal{N}_1 = n_1$ at time $t = t_1$. It seems natural to assume an intermediate state described by the random variable $\mathcal{N}_2 = n_2$ at $t = t_2$ with the implicit order in time: $t_1 \geq t_2 \geq t_3$ (figure 3.4). The value of the variable $\mathcal{N}_2$, however, need not be unique or in other words, there may be several paths or trajectories leading from (n_3, t_3) to (n_1, t_1). Since we are interested in the propagation of a distribution and not in a single trajectory the probability distribution at intermediate times is important. Therefore the individual values of the random variables are replaced by probabilities, $\mathcal{N} = n \implies P(\mathcal{N} = n, t) = P(n, t)$,[16] and an equation is obtained that encapsulates the full diversity of different sources of randomness with the exception of quantum uncertainty. The only restriction in the generally used form of this equation is the Markov property of the stochastic process. This general equation is named *Chapman-Kolmogorov equation* after the British geophysicist and mathematician Sydney Chapman and the Russian mathematician Andrey Kolmogorov and in the rest of this section we shall be concerned with it in its various forms.

3.2.1 Forward Chapman-Kolmogorov equation

A forward equation predicts the future of a system from given information of the present state, and this is the most common strategy in modeling dynamical phenomena. It allows for direct comparison with experimental data, which in observations are, of course, also recorded in the forward direction. There are however problems like the computation of first passage times or the reconstruction of phylogenetic trees that call for an opposite strategy, which aims at a reconstruction of the past from present day information. In such cases so-called backward equations facilitate the analysis (see, e.g., section 3.3).

3.2.1.1 Discrete and continuous Chapman-Kolmogorov equations

The relation between the three random variables $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ can be illustrated by application of set theoretical considerations. Let A, B and C be the corresponding events and $B_k\,(k = 1, \ldots, n)$ a partition of B into n mutually exclusive *subevents*. Then, if all events of one kind are included in the summation the corresponding variable $\mathcal{B}$ is eliminated:

[16] Here, we need not yet specify whether the sample space is discrete, $P_n(t)$, or continuous, $P(x, t)$, but we shall do so in section 3.2.1.1.

$$\sum_k P(A \cap B_k \cap C) \;=\; P(A \cap C) \;.$$

The relation can be easily verified by means of Venn diagrams. Translating this results into the language of stochastic process we assume first to be dealing with a discrete state space and accordingly the random variables $\mathcal{N} \in \mathbb{N}$ are defined on the integers. Then we can simply make use of state space covering and find for the marginal probability

$$P(n_1, t_1) \;=\; \sum_{n_2} P(n_1, t_1; n_2, t_2) \;=\; \sum_{n_2} P(n_1, t_1 |\, n_2, t_2)\, P(n_2, t_2) \;.$$

Next we introduce a third event (n_3, t_3) (figure 3.4) and describe the process by the equations for conditional probabilities

$$\begin{aligned} P(n_1, t_1 |\, n_3, t_3) \;&=\; \sum_{n_2} P(n_1, t_1; n_2, t_2 |\, n_3, t_3) \;= \\ &=\; \sum_{n_2} P(n_1, t_1 |\, n_2, t_2; n_3, t_3)\, P(n_2, t_2 |\, n_3, t_3) \;. \end{aligned}$$

Both equations are of general validity for all stochastic processes, and the series could be extended further to four, five events and so on. Finally, adopting the Markov assumtion and introducing the time order $t_1 \geq t_2 \geq t_3$ provides the basis for dropping the dependence on (n_2, t_2) in the doubly conditioned probability and leads to

$$P(n_1, t_1 | n_3, t_3) \;=\; \sum_{n_2} P(n_1, t_1 | n_2, t_2)\, P(n_2, t_2 | n_3, t_3) \;. \tag{3.36}$$

This is the *Chapman-Kolmogorov equation* in its simplest general form. Equation (3.36) can be interpreted as a matrix multiplication, $\mathrm{C} = \mathrm{A} \cdot \mathrm{B}$ with $c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}$, where the eliminated dimension of the matrices, m, reflects on the size of the event space of the eliminated variable n_2, it could be even countably infinite.

The extension from the discrete case to probability densities is straightforward. By the same token we find for the continuous case

$$p\,(x_1, t_1) \;=\; \int \mathrm{d}x_2\, p\,(x_1, t_1; x_2, t_2) \;=\; \int \mathrm{d}x_2\, p\,(x_1, t_1 | x_2, t_2)\, p\,(x_2, t_2) \;,$$

and the extension to three events leads to

$$\begin{aligned} p\,(x_1, t_1 | x_3, t_3) \;&=\; \int \mathrm{d}x_2\, p\,(x_1, t_1; x_2, t_2 | x_3, t_3) \;= \\ &=\; \int \mathrm{d}x_2\, p\,(x_1, t_1 | x_2, t_2; x_3, t_3)\, p\,(x_2, t_2 | x_3, t_3) \;. \end{aligned}$$

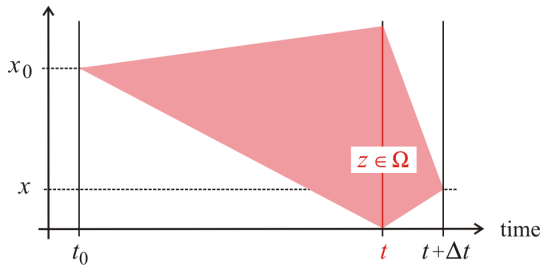


Fig. 3.6 Time order in the differential Chapman-Kolmogorov equation (dCKE) **.** The one-dimensional sketch shows the notation used in the derivation of the forward dCKE. The variable $\mathbf{z}$ is integrated over the entire sample space Ω in order to sum up all trajectories leading from $(\mathbf{x}_0, t_0)$ via $(\mathbf{z}, t)$ to $(\mathbf{x}, t + \Delta t)$.

For $t_1 \geq t_2 \geq t_3$ and making again use of the Markov property we obtain the continuous version of the Chapman-Kolmogorov equation:

$$p\,(x_1, t_1|x_3, t_3) \;=\; \int \mathrm{d}x_2 p\,(x_1, t_1|x_2, t_2)\, p\,(x_2, t_2|x_3, t_3)\;. \tag{3.37}$$

Equation (3.37) is of very general nature. The only relevant approximation is the assumption of a Markov process, which is empirically well justified in most applications in physics, chemistry and biology. General validity is commonly accompanied by a variety of different solutions and the Chapman-Kolmogorov equation is no exception in this aspect. The generality of (3.37) in the description of a stochastic process becomes evident when the evolution in time is continued $t_1 \geq t_2 \geq t_3 \geq t_4 \geq t_5 \ldots$, and complete summations over all intermediate states are performed

$$p\,(x_1, t_1|x_n, t_n) \;=\; \int \mathrm{d}x_2 \cdots \int \mathrm{d}x_{n-1} p\,(x_1, t_1|x_2, t_2) \ldots p\,(x_{n-1}, t_{n-1}|x_n, t_n)\,.$$

Sometimes it is useful – and we shall adopt also this notation here – to indicate an initial state by the doublet (x_0, t_0) and apply physical notation of time.

3.2.1.2 Differential Chapman-Kolmogorov forward equation

Since we aim at a description of processes the Chapman-Kolmogorov equations in discrete and continuous form as expressed in equations (3.36) and (3.37), respectively, provide a general definition of Markov processes but they are not really useful to describe the temporal evolution. Much better suited for describing stochastic processes as well as analyzing nature and properties of solutions or performing actual calculations are equations in differential form. In a way the differential formulation of basic stochastic processes can

be compared to the invention of calculus by Gottfried Wilhelm Leibniz and Isaac Newton, which provides the ultimate basis for all modeling by means of differential equations. Analytical solution or numerical integration of such a *differential Chapman-Kolmogorov equation* (dCKE) is then expected to provide the desired description of processes. A differential form of the Chapman-Kolmogorov equation has been derived by Crispin Gardiner [157, pp. 48-51].[17] We shall follow here, in essence, a simpler approach given more recently by Mukhtar Ullah and Olaf Wolkenhauer [434, 435].

The Chapman-Kolmogorov equation is considered for an interval $t \to t+\Delta t$ defined for a sample space Ω and the initial conditions $(\mathbf{x}_0, t_0)$:

$$p\,(\mathbf{x}, t+\Delta t|\,\mathbf{x}_0, t_0) \;=\; \int_\Omega \mathrm{d}\mathbf{z}\; p\,(\mathbf{z}, t+\Delta t|\,\mathbf{x}_0, t_0)\, p\,(\mathbf{z}, t|\,\mathbf{x}_0, t_0)\,, \qquad (3.37')$$

whereby the consistency equation (3.25) is fulfilled. As illustrated in figure 3.6 the probability of the transition from $(\mathbf{x}_0, t_0)$ to $(\mathbf{x}, t+\Delta t)$ is obtained by integration over all probabilities to occur via an intermediate, $(\mathbf{x}_0, t_0) \to (\mathbf{z}, t) \to (\mathbf{x}, t+\Delta t)$. In order to simplify derivation and notation we shall assume fixed sharp initial conditions $(\mathbf{x}_0, t_0)$ or in other words, the unconditioned probability of the state $(\mathbf{x}, t)$ is the same as the probability of the transition from $(\mathbf{x}_0, t_0) \to (\mathbf{x}, t)$:

$$p\,(\mathbf{x}, t) \;=\; p\,(\mathbf{x}, t|\,\mathbf{x}_0, t_0) \;\text{ with }\; p\,(\mathbf{x}, t_0) \;=\; \delta(\mathbf{x}-\mathbf{x}_0). \qquad (3.38)$$

We introduce the time derivative by tacitly assuming that the probability $p\,(\mathbf{x}, t)$ is differentiable with respect to time:

$$\frac{\partial}{\partial t}\, p\,(\mathbf{x}, t) \;=\; \lim_{\Delta t \to 0} \frac{1}{\Delta t}\Big(p\,(\mathbf{x}, t+\Delta t) \;-\; p\,(\mathbf{x}, t)\Big) \qquad (3.39)$$

Introducing the CKE in form (3.37') and multiplying $p\,(\mathbf{x}, t)$ formally by one in the form of the normalization condition of probabilities,[18]

$$1 \;=\; \int_\Omega \mathrm{d}\mathbf{z}\, p\,(\mathbf{z}, t+\Delta t|\,\mathbf{x}, t)\;,$$

we can rewrite equation (3.39) as

$$\begin{aligned} \frac{\partial}{\partial t}\, p\,(\mathbf{x}, t) \;=\; \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_\Omega \mathrm{d}\mathbf{z} \Big(& p\,(\mathbf{x}, t+\Delta t|\,\mathbf{z}, t)\, p\,(\mathbf{z}, t) \;- \\ & -\; p\,(\mathbf{z}, t+\Delta t|\,\mathbf{x}, t)\, p\,(\mathbf{x}, t)\Big)\;. \end{aligned} \qquad (3.40)$$

[17] The derivation is contained already in the first edition of Gardiner's *Handbook of stochastic methods* [156] and it has been Crispin Gardiner who coined the name *differential Chapman-Kolmogorov equation*.

[18] It is important to note a useful trick in the derivation: By the substitution the time order is reversed in the integral.

For the purpose of integration the sample space Ω is divided up into to parts with respect to an arbitrarily small parameter $\epsilon > 0$: $\Omega = I_1 + I_2$. Using the notion of continuity (section 3.1.3.5) the region I_1 defined by $\|\mathbf{x} - \mathbf{z}\| < \epsilon$ represents a continuous process.[19] In the second part of the sample space Ω, I_2 with $\|\mathbf{x}-\mathbf{z}\| \geq \epsilon$, the norm cannot become arbitrarily small corresponding to a jump process. For the derivative taken on entire sample space Ω we get:

$$\begin{aligned}
\frac{\partial}{\partial t} p\,(\mathbf{x}, t) \;&=\; I_1 \,+\, I_2\,, \quad \text{with} \\
I_1 \;&=\; \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\|\mathbf{x}-\mathbf{z}\|<\epsilon} \mathrm{d}\mathbf{z} \Big(p\,(\mathbf{x}, t+\Delta t|\, \mathbf{z}, t)\, p\,(\mathbf{z}, t) - \\
&\qquad\qquad - \, p\,(\mathbf{z}, t+\Delta t|\, \mathbf{x}, t)\, p\,(\mathbf{x}, t) \Big)\,, \quad \text{and} \qquad (3.41) \\
I_2 \;&=\; \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\|\mathbf{x}-\mathbf{z}\|\geq\epsilon} \mathrm{d}\mathbf{z} \Big(p\,(\mathbf{x}, t+\Delta t|\, \mathbf{z}, t)\, p\,(\mathbf{z}, t) - \\
&\qquad\qquad - \, p\,(\mathbf{z}, t+\Delta t|\, \mathbf{x}, t)\, p\,(\mathbf{x}, t) \Big)\,.
\end{aligned}$$

In the first region I_1 with $\|\mathbf{x} - \mathbf{z}\| < \epsilon$ we introduce $\mathbf{r} = \mathbf{x} - \mathbf{z}$

$$\begin{aligned}
I_1 \;&=\; \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\|\mathbf{r}\|<\epsilon} \mathrm{d}\mathbf{r} \Big(p\,(\mathbf{x}, t+\Delta t|\, \mathbf{x}-\mathbf{r}, t)\, p\,(\mathbf{x}-\mathbf{r}, t) - \\
&\qquad\qquad - \, p\,(\mathbf{x}-\mathbf{r}, t+\Delta t|\, \mathbf{x}, t)\, p\,(\mathbf{x}, t) \Big)\,, \quad \text{and define}
\end{aligned}$$

$$f(\mathbf{x}, \mathbf{r}) \;\doteq\; p\,(\mathbf{x}+\mathbf{r}, t+\Delta t|\mathbf{x}, t)\, p\,(\mathbf{x}, t) \quad \text{yielding}$$

$$I_1 \;=\; \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\|\mathbf{r}\|<\epsilon} \mathrm{d}\mathbf{r} \Big(f(\mathbf{x}-\mathbf{r}, \mathbf{r}) - f(\mathbf{x}, -\mathbf{r}) \Big) \;=\; \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{\|\mathbf{r}\|<\epsilon} \mathrm{d}\mathbf{r} F(\mathbf{x}, \mathbf{r}).$$

Next the integrand is expanded in a Taylor series in $-\mathbf{r}$ about $\mathbf{r} = 0$:

$$F(\mathbf{x}, \mathbf{r}) = -f(\mathbf{x}, -\mathbf{r}) + f(\mathbf{x}, \mathbf{r}) + \sum_i (-r_i) \frac{\partial f(\mathbf{x}, \mathbf{r})}{\partial x_i} + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f(\mathbf{x}, \mathbf{r})}{\partial x_i \partial x_j} + \ldots$$

In this Taylor expansion the constant term is zero and all terms higher than second order vanish in the limit $\Delta t \to 0$ [157, 434]. Provided the differentiability conditions are fulfilled we obtain in the limit $\epsilon \to 0$:

$$I_1 \;=\; -\sum_i \frac{\partial}{\partial x_i} \Big(A_i(\mathbf{x}, t)\, p\,(\mathbf{x}, t) \Big) \,+\, \frac{1}{2} \sum_i \sum_j \frac{\partial^2}{\partial x_i \partial x_j} \Big(B_{ij}(\mathbf{x}.t)\, p\,(\mathbf{x}, t) \Big)\,. \qquad (3.42)$$

which defines a Fokker-Planck equation. We remark that in the limit $\epsilon \to 0$ the continuous part of the process becomes equivalent to an equation for the differential increments of the random vector $\vec{\mathcal{X}}(t)$ describing a single trajec-

[19] The notation $\|\cdot\|$ refers to a suitable vector norm, here the L^1 norm, $\|\mathbf{y}\| = \sum_k |y_k|$. In the one-dimensional case we would just use the absolute value $|y|$.

tory:

$$\vec{\mathcal{X}}(t+\mathrm{dt}) \;=\; \vec{\mathcal{X}}(t) \;+\; \mathbf{A}\big(\vec{X}(t),t\big)\;\mathrm{dt} \;+\; \Big(\mathrm{B}\big(\vec{\mathcal{X}}(t),t\big)\;\mathrm{dt}\Big)^{\frac{1}{2}}\;. \tag{3.43}$$

In the terminology used in physics $\mathbf{A}$ is the *drift vector* and B is the *diffusion matrix* of the stochastic process. In other words, for $\epsilon \to 0$ and continuity of the process the expectation value of the increment vector expressed by $\vec{\mathcal{X}}(t+\mathrm{dt}) - \vec{\mathcal{X}}(t)$ approaches $\mathbf{A}\big(\vec{X}(t),t\big)\,\mathrm{dt}$ and its covariance converges to $\mathrm{B}\big(\vec{X}(t),t\big)\,\mathrm{dt}$. Equation (3.43) is a stochastic differential equation or Langevin equation (see section 3.4.1).
The second part of the integration over sample space Ω involves the probability rate for jumps:

$$\begin{aligned} I_2 \;=\; \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{\|\mathbf{x}-\mathbf{z}\|\geq\epsilon}\mathrm{d}\mathbf{z}\Big(&p\left(\mathbf{x},t+\Delta t|\,\mathbf{z},t\right)p\left(\mathbf{z},t\right) - \\ &-\; p\left(\mathbf{z},t+\Delta t|\,\mathbf{x},t\right)p\left(\mathbf{x},t\right)\Big)\;. \end{aligned}$$

The condition for a jump process is $\|\mathbf{x}-\mathbf{z}\| \geq \epsilon$ (section 3.1.3.5) and accordingly we have

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\left(p\left(\mathbf{x},t+\Delta t|\,\mathbf{z},t\right)p\left(\mathbf{z},t\right)\right) \;=\; W(\mathbf{x}|\,\mathbf{z},t)\,p\left(\mathbf{z},t\right)\;, \tag{3.44}$$

where $W(\mathbf{x}|\,\mathbf{z},t)$ is the transition rate for the jump $\mathbf{z}\to\mathbf{x}$. By the same token we define a transition rate for the jump in the reverse direction $\mathbf{x}\to\mathbf{z}$. As $\epsilon\to 0$ the integration is extended over the whole same space Ω and eventually we obtain

$$\lim_{\epsilon\to 0} I_2 \;=\; \int_{\Omega}\mathrm{d}\mathbf{z}\Big(W(\mathbf{x}|\,\mathbf{z},t)\,p\left(\mathbf{z},t\right) \;-\; W(\mathbf{z}|\,\mathbf{z},t)\,p\left(\mathbf{x},t\right)\Big)\;, \tag{3.45}$$

which completes the somewhat simplified derivation of the differential Chapman-Kolmogorov equation.

The evolution of the system is now expressed in terms of functions $\mathbf{A}(\mathbf{x},t)$, which correspond to the functional relations in conventional differential equations, a diffusion matrix $\mathrm{B}(\mathbf{x},t)$, and transition matrix for discontinuous jumps $\mathrm{W}(\mathbf{x}|\mathbf{z},t)$:

$$\frac{\partial p\,(\mathbf{x},t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}\Big(A_i(\mathbf{x},t)\,p\,(\mathbf{x},t)\Big) + \tag{3.46a}$$

$$+\frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}\Big(B_{ij}(\mathbf{x},t)\,p\,(\mathbf{x},t)\Big) + \tag{3.46b}$$

$$+\int \mathrm{d}\mathbf{z}\,\Big(\mathrm{W}(\mathbf{x}\,|\mathbf{z},t)\,p\,(\mathbf{z},t) - \mathrm{W}(\mathbf{z}\,|\mathbf{x},t)\,p\,(\mathbf{x},t)\Big). \tag{3.46c}$$

Equation (3.46) is called a *forward equation* in the sense of figure 3.22. Surface terms at the boundary of the domain of $\mathbf{x}$ have been neglected in the derivation [157, p. 50]. This assumption is not critical for most cases considered here. It is always correct for infinite domains because the probabilities vanish in the limit $\lim_{\mathbf{x}\to\pm\infty} p\,(\mathbf{x},t) = 0$.

From a mathematical purist's point of view it is not clear from the derivation that solutions of the differential Chapman-Kolmogorov equation (3.46) exist, are unique and are solutions to the Chapman-Kolmogorov equation (3.37) as well. It is true, however, that the set of conditional probabilities obeying equation (3.46) does generate a Markov process in the sense that the joint probabilities produced satisfy all probability axioms. It has been shown, however, that a non-negative solution to the differential Chapman-Kolmogorov equations exists and satisfies the Chapman-Kolmogorov equation under certain conditions (see [165, Vol.II]):

(1) $\mathbf{A}(\mathbf{x},t) = \{A_i(\mathbf{x},t);\, i = 1,\ldots\}$ and $\mathrm{B}(\mathbf{x},t) = \{B_{ij}(\mathbf{x},t);\, i,j = 1,\ldots\}$ are vectors and positive semidefinite matrices[20] of functions, respectively,
(2) $\mathrm{W}(\mathbf{x}|\mathbf{z},t)$ and $\mathrm{W}(\mathbf{z}|\mathbf{x},t)$ are non-negative quantities,
(3) the initial condition has to satisfy $p\,(\mathbf{x},t_0|\mathbf{x}_0,t_0) = \delta(\mathbf{x}_0 - \mathbf{x})$, and
(4) appropriate boundary conditions have to be fulfilled.

The boundary conditions are very hard to specify for the full equation but can be discussed precisely for special cases, for example in the case of the Fokker-Planck equation [383].

The nature of the different stochastic processes associated with the three terms in equation (3.46), $\mathbf{A}(\mathbf{x},t)$, $\mathrm{B}(\mathbf{x},t)$, $\mathrm{W}(\mathbf{x}\,|\,\mathbf{z},t)$ and $\mathrm{W}(\mathbf{z}\,|\,\mathbf{x},t)$, is visualized by setting some parameters equal to zero and analyzing the remaining equation. We shall discuss here four cases that are modeled by different equations (for relations between them see figure 3.1).

(i) $\mathrm{B} = 0$, $\mathrm{W} = 0$, deterministic drift process: Liouville equation,
(ii) $\mathbf{A} = 0$, $\mathrm{W} = 0$, drift free diffusion or Wiener process: diffusion equation,
(iii) $\mathrm{W} = 0$, drift and diffusion process: Fokker-Planck equation, and
(iv) $\mathbf{A} = 0$, $\mathrm{B} = 0$, pure jump process: master equation.

[20] A positive definite matrix has exclusively positive eigenvalues, $\lambda_k > 0$ whereas a positive semidefinite matrix has non-negative eigenvalues, $\lambda_k \geq 0$.

The first term in differential Chapman-Kolmogorov equation, equation (3.46a) is the probabilistic version of a differential equation describing deterministic motion, which is known as *Liouville equation* named after the French mathematician Joseph Liouville. It is a fundamental equation of statistical mechanics and will be discussed in some detail subsection 3.2.2.1. With respect to the theory of stochastic processes (3.46a) it encapsulates the drift of a probability distribution.

The second term in equation (3.46) describes spreading of probability densities by diffusion and is called a stochastic *diffusion equation*. In pure form it is represented by the Wiener process, which got the name from the American mathematician Norbert Wiener and which can be understood as the continuous time and continuous space limit of the one-dimensional random walk (see figure 3.3). The Wiener process is fundamental for understanding stochasticity in continuous space and time and will be discussed in subsection 3.2.2.2.

Combining equations (3.46a) and (3.46b) yields the Fokker-Planck equation, which we repeat here because of its general importance:

$$\frac{\partial p\left(\mathbf{x},t\right)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}\Big(A_i(\mathbf{x},t)\,p\left(\mathbf{x},t\right)\Big) + \frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i \partial x_j}\Big(B_{ij}(\mathbf{x},t)\,p\left(\mathbf{x},t\right)\Big)\,. \quad (3.47)$$

The equation is named after two physicists, the Dutchman Adriaan Daniël Fokker and the German Max Planck. Fokker-Planck equations are frequently used in physics to model and analyze processes with fluctuations [383].

If only the third term of the differential Chapman-Kolmogorov equation, (3.46c), has nonzero elements, the variables $\mathbf{x}$ and $\mathbf{z}$ change only in steps and the corresponding differential equation is called a *master equation*. Master equations are the most important tools for describing processes in discrete spaces, $\mathcal{X}(t) \in \mathbb{N}$. We shall discuss specific examples in sections 3.2.2.4 and 3.2.4 and treat them in a whole section (section 3.2.3). In particular, master equations are indispensable for modeling chemical reactions or biological processes with small particle numbers. Specific applications in chemistry and biology will be presented in two separate chapters 4 and 5.

It is important to stress that the mathematical expressions of the three contributions to the general stochastic process represent a pure formalism that can be applied equally well to problems in physics, chemistry, biology, sociology, economics or other disciplines. Specific empirical knowledge enters the model in form of the parameters: the drift vector $\mathbf{A}$, the diffusion matrix B, and the jump transition matrix W. By means of examples we shall show how physical laws are encapsulated in regularities among the parameters.

3.2.2 Examples of stochastic processes

In this section we present examples of stochastic processes with characteristic properties that will be useful as references in the forthcoming applications:

(i) the Liouville process, (ii) the Wiener process, (iii) the Ornstein-Uhlenbeck process, (iv) the Poisson process, and (v) the random walk in one dimension.

3.2.2.1 Liouville process

The Liouville equation[21] is the straightforward link between deterministic motion and stochastic processes. As indicated in figure 3.1 all elements of the jump transition matrix W and the diffusion matrix B are zero and what remains is a differential equation falling into the class of *Liouville equations* from classical mechanics. A Liouville equation is used commonly for the description of the deterministic motion of particles in *phase space*.[22] Following [157, p. 54] we show that deterministic trajectories are identical to solutions of the differential Chapman-Kolmogorov equation with D = 0 and W = 0 and then relate the result to Liouville's theorem in classical mechanics [292, 293].

First we consider deterministic motion as described by the differential equation

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{\xi}(t)}{\mathrm{dt}} &= \mathbf{A}\big(\boldsymbol{\xi}(t),t\big) \quad \text{with} \quad \boldsymbol{\xi}(t_0) = \boldsymbol{\xi}_0 \quad \text{and} \\ \boldsymbol{\xi}(t) &= \boldsymbol{\xi}_0 + \int_{t_0}^{t} \mathrm{d}\boldsymbol{\xi}\, \mathbf{A}\big(\boldsymbol{\xi}(t),t\big)\,, \end{aligned} \tag{3.48}$$

that can be understood as a kind of degenerate Markov process in which the probability distribution degenerates to a Dirac delta-function $p\,(\mathbf{x},t) = \delta\big(\mathbf{x} - \boldsymbol{\xi}(t)\big)$.[23] We may relax the initial conditions $\boldsymbol{\xi}(t_0) = \boldsymbol{\xi}_0$ or $p\,(\mathbf{x},t_0) = \delta(\mathbf{x}-\boldsymbol{\xi}_0)$ to $p\,(\mathbf{x},t_0) = p\,(\mathbf{x}_0)$ and then the result is a distribution migrating through space with unchanged shape (figure 3.7) instead of a delta function travelling on a single trajectory (see equation (3.53') below).

By setting B = 0 and W = 0 in the dCKE we obtain for the Liouville process

$$\frac{\partial p\,(\mathbf{x},t)}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}\Big(A_i(\mathbf{x},t)\, p\,(\mathbf{x},t)\Big)\,, \tag{3.49}$$

The goal is now to show equivalence to the differential equation (3.48) in form of the common solution

$$p\,(\mathbf{x},t) = \delta\big(\mathbf{x} - \boldsymbol{\xi}(t)\big)\,. \tag{3.50}$$

The proof is done by direct substitution

---

[21] The notion Liouville equation has been created by Josiah Willard Gibbs [162].

[22] Phase space is an abstract space, which is particularly useful for the visualization of particle motion. The six independent coordinates of particle S_k are the position coordinates $\mathbf{q}_k = (q_{k1}, q_{k2}, q_{k3})$ and the (linear) momentum coordinates $\mathbf{p}_k = (p_{k1}, p_{k2}, p_{k3})$. In Cartesian coordinates they read: $\mathbf{q}_k = (x_k, y_k, z_k)$ and $\mathbf{p}_k = m_k \cdot \mathbf{v}_k$ with $\mathbf{v} = (v_x, v_y, v_z)$ being the velocity vector.

[23] For simplicity we write $p\,(\mathbf{x},t)$ instead of the conditional probability $p\,(\mathbf{x},t|\,\mathbf{x}_0,t_0)$ as long as the initial condition $(\mathbf{x}_0,t_0)$ refer to the sharp density $p\,(\mathbf{x},t_0) = \delta(\mathbf{x}-\mathbf{x}_0)$.

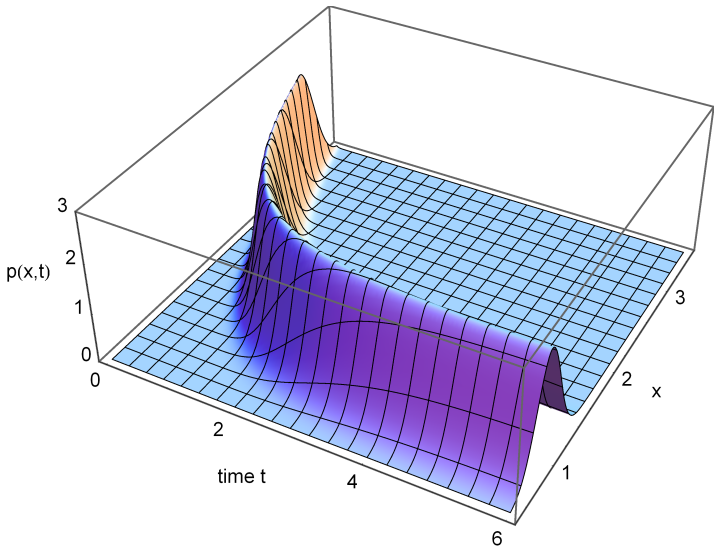


Fig. 3.7 Probability density of a Liouville process. The figure shows the migration of a normal distribution $p\,(x) = \sqrt{k/(\pi s^2)}\exp\bigl(-k(x-\mu)^2/s^2\bigr)$ along a trajectory corresponding to the expectation value of an Ornstein-Uhlenbeck process: $\xi(t) = \mu + (\xi_0 - \mu)\exp\,(-kt)$ (section 3.2.2.3). The expression for the density is

$$p\,(x,t) = \sqrt{\tfrac{k}{\pi\,s^2}}\cdot e^{-k\left(x-\mu-(\xi_0-\mu)\,\exp\,(-kt)\right)/s^2},$$

and the long-time limit of the distribution, $p\,(\bar{x})$, is a normal distribution with mean $\mathrm{E}(\bar{x}) = \mu$ and variance $\mathrm{var}(\bar{x}) = \sigma^2 = s^2/2k$. Choice of parameters: $\xi_0 = 3\,[l.u.]$, $k = 1\,[t.u.]^{-1}$, $\mu = 1\,[l.u.]$, $s = 1/4\,[t.u.]^{-1/2}$.

$$-\sum_i \frac{\partial}{\partial x_i}\Bigl(A_i(\mathbf{x},t)\,\delta\bigl(\mathbf{x}-\boldsymbol{\xi}(t)\bigr)\Bigr) \;=\; -\sum_i \frac{\partial}{\partial x_i}\Bigl(A_i\bigl(\boldsymbol{\xi}(t),t\bigr)\;\delta\bigl(\mathbf{x}-\boldsymbol{\xi}(t)\bigr)\Bigr) \;=$$

$$= \;-\sum_i\Bigl(A_i\bigl(\boldsymbol{\xi}(t),t\bigr)\,\frac{\partial}{\partial x_i}\;\delta\bigl(\mathbf{x}-\boldsymbol{\xi}(t)\bigr)\Bigr)\,,$$

since $\boldsymbol{\xi}$ does not depend on $\mathbf{x}$ and

$$\frac{\partial p\,(\mathbf{x},t)}{\partial t} \;=\; \frac{\partial}{\partial t}\;\delta\bigl(\mathbf{x}-\boldsymbol{\xi}(t)\bigr) \;=\; -\sum_i\Bigl(\frac{\mathrm{d}\xi_i(t)}{\mathrm{dt}}\cdot\frac{\partial}{\partial x_i}\;\delta\bigl(\mathbf{x}-\boldsymbol{\xi}(t)\bigr)\Bigr)\,.$$

Making use of equation (3.48) in component form

$$\frac{\mathrm{d}\xi_i(t)}{\mathrm{dt}} \;=\; A_i\bigl(\boldsymbol{\xi}(t),t\bigr)$$

we see that the sums in the expressions on the last two lines are equal. □

The following part on Liouville's equation illustrates how empirical science – here Newtonian mechanics – enters a formal stochastic equation. In

Hamiltonian mechanics [190, 191] dynamical systems may be represented by a *density function* or *classical density matrix* $\varrho(\mathbf{q},\mathbf{p})$ in phase space. The density function allows for the calculation of system properties. Commonly it is normalized such that the expected total number of particles is the integral over phase space:

$$N \;=\; \int\cdots\int \varrho(\mathbf{q},\mathbf{p})\,(\mathrm{d}q)^n(\mathrm{d}p)^n \;.$$

The evolution of the system is described by a time dependent density that is commonly denoted as $\varrho\left(\mathbf{q}(t),\mathbf{p}(t),t\right)$ with the initial conditions $\varrho\left(\mathbf{q}_0,\mathbf{p}_0,t_0\right)$. For a single particle S_k the generalized spatial coordinated q_{ki} are related to conjugate momenta p_{ki} by Newton's equations of motion

$$\frac{\mathrm{d}p_{ki}}{\mathrm{dt}} \;=\; f_{ki}(\mathbf{q}) \quad\text{and}\quad \frac{\mathrm{d}q_{ki}}{\mathrm{dt}} \;=\; \frac{1}{m_k}\,p_{ki};\quad i=1,2,3\,;\; k=1,\ldots,n\;, \tag{3.51}$$

where f_{ki} is the component of the force acting on particle S_k in the direction of q_{ki} and m_k the particle mass, respectively. Liouville's theorem based on Hamiltonian mechanics of an n particle system makes a statement on the evolution of the density ϱ

$$\frac{\mathrm{d}\varrho(\mathbf{q},\mathbf{p},t)}{\mathrm{dt}} \;=\; \frac{\partial\varrho}{\partial t} + \sum_{k=1}^{n}\sum_{i=1}^{3}\left(\frac{\partial\varrho}{\partial q_{ki}}\,\frac{\mathrm{d}q_{ki}}{\mathrm{dt}} + \frac{\partial\varrho}{\partial p_{ki}}\,\frac{\mathrm{d}p_{ki}}{\mathrm{dt}}\right) \;=\; 0\,: \tag{3.52}$$

The density function does not change with time, it is a constant of the motion and therefore constant along the trajectory in phase space.

Now, we can show that equation (3.52) can be transformed into a Liouville equation (3.49). We insert the individual time derivatives and find

$$\frac{\partial\varrho(\mathbf{q},\mathbf{p},t)}{\partial t} \;=\; -\sum_{k=1}^{n}\sum_{i=1}^{3}\left(\frac{1}{m_i}p_{ki}\,\frac{\partial}{\partial q_{ki}}\varrho(\mathbf{q},\mathbf{p},t) \;+\; f_{ki}\,\frac{\partial}{\partial p_{ki}}\varrho(\mathbf{q},\mathbf{p},t)\right). \tag{3.53}$$

Equation (3.53) has already the form of a differential Chapman-Kolmogorov equation (3.49) with $\mathrm{B}=0$ and $\mathrm{W}=0$ as follows from

$$\begin{aligned}
\varrho(\mathbf{q},\mathbf{p},t) \;&\equiv\; p(\mathbf{x},t)\,,\\
\mathbf{x} \;&\equiv\; (q_{11},\ldots,q_{n3},p_{11},\ldots,p_{n3}) \quad\text{and}\\
\mathbf{A} \;&\equiv\; \left(\frac{1}{m_1}p_{11},\ldots,\frac{1}{m_n}p_{n3},f_{11},\ldots,f_{n3}\right)\;.
\end{aligned}$$

where the $6n$ components of $\mathbf{x}$ representing $3n$ coordinates for the positions and $3n$ coordinates for the linear momenta of n particles. Finally, we indicate the relations of the probability density $p(\mathbf{x},t)$ to equations (3.48) and (3.49): The density function is the expectation value of the probability distribution,

$$\mathbf{x}(t) \;=\; \big(\mathbf{q}(t), \mathbf{p}\,(t)\big) \;=\; \mathrm{E}\Big(\varrho\big(\mathbf{q}(t), \mathbf{p}\,(t), t\big)\Big)\;, \tag{3.54}$$

and it fulfils the Liouville ODE as well as the Chapman-Kolmogorov equation:

$$\begin{aligned}
\frac{\partial p\,(\mathbf{x}, t)}{\partial t} \;&\equiv\; \frac{\partial \varrho\,(\mathbf{q}, \mathbf{p}, t)}{\partial t} \;= \\
&=\; -\sum_{i=1}^{3n} \left(\frac{1}{m_i} p_i \,\frac{\partial}{\partial q_i} \varrho\,(\mathbf{q}, \mathbf{p}, t) \;+\; f_i \,\frac{\partial}{\partial p_i} \varrho\,(\mathbf{q}, \mathbf{p}, t) \right) \;= \\
&=\; -\sum_{i=1}^{6n} \frac{\partial}{\partial x_i} \Big(A_i(\mathbf{x}, t)\, p\,(\mathbf{x}, t) \quad \text{and} \qquad (3.53')
\end{aligned}$$

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{dt}} \;=\; \mathbf{A}\big(\mathbf{x}(t), t\big)\;. \tag{3.51'}$$

In other words, the Liouville equation states the conservation of the density matrix $\varrho\,(\mathbf{q}, \mathbf{p}, t)$ in phase space as shown for a normal density in figure (3.7).

3.2.2.2 Wiener process

The Wiener process named after the American mathematician and logician Norbert Wiener is fundamental in many aspects. It is often used as a synonym for Brownian motion or *white noise* and describes among other things the diffusion due to random fluctuations caused by thermal motion. The fluctuation driven random variable is denoted by $\mathcal{W}(t)$ and it is characterized by the cumulative probability distribution,

$$P\,(\mathcal{W}(t) \le w) = \int_{-\infty}^{w} p\,(u, t)\,\mathrm{d}u\;,$$

where $p\,(u, t)$ still has to be determined. From the point of view of stochastic processes the probability density of the Wiener process is the solution of the differential Chapman-Kolmogorov equation in one variable with a diffusion term $\mathrm{B} = D = 1$, zero drift, $\mathbf{A} = 0$, and no jumps, $\mathrm{W} = 0$:

$$\frac{\partial p\,(w, t)}{\partial t} \;=\; \frac{1}{2}\,\frac{\partial^2}{\partial w^2}\, p\,(w, t) \quad \text{with} \quad p\,(w, t_0) = \delta(w - w_0)\;. \tag{3.55}$$

Again a sharp initial condition (w_0, t_0) is assumed and we write for short $p\,(w, t) = p\,(w, t|w_0, t_0)$.

The closely related equation

$$\frac{\partial\, c\,(x, t)}{\partial t} \;=\; D\,\frac{\partial^2}{\partial x^2}\, c(x, t) \quad \text{with} \quad c(x, t_0) \;=\; c_0(x)\;, \tag{3.56}$$

is called *diffusion equation*, because $c(x,t)$ describes the spreading of concentrations in homogeneous media driven by thermal molecular motion (for a detailed mathematical description of diffusion see, for example, [79, 174]). The parameter D is called the diffusion coefficient and here it is assumed to be a constant that means independent of space and time. The one-dimensional version of (3.56) is identical with the equation (3.55) with $D = 1/2$.[24] The three-dimensional version of equation (3.55) occurs in physics and chemistry in connection with particle numbers or concentrations as functions of 3D-space and time, $c(\mathbf{r},t)$, which fulfill

$$\frac{\partial\, c\,(\mathbf{r},t)}{\partial t} \;=\; D\,\nabla^2 c\,(\mathbf{r},t) \;\;\text{with}\;\; \mathbf{r} = (x,y,z)\,,\; \nabla^2 = \Big(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\Big)\,, \quad (3.57)$$

and the initial condition $c\,(\mathbf{r},t_0) = c_0(\mathbf{r})$. The diffusion equation has been derived first by Adolf Fick in 1855 [367]. Replacing the concentration by the temperature distribution in an one-dimensional object $c(x,t) \Leftrightarrow u(x,t)$ and the diffusion constant by the thermal diffusivity, $D \Leftrightarrow \alpha$, the diffusion equation (3.56) becomes the heat equation, which describes the time dependence of the distribution of heat over a given region time.

Solutions of equation (3.55) can be calculated readily by means of the characteristic function:

$$\phi(s,t) \;=\; \int_{-\infty}^{+\infty} \mathrm{d}w\, p\,(w,t)\, e^{\imath s\, w} \;\;\text{and}$$

$$\frac{\partial\phi(s,t)}{\partial t} \;=\; \int_{-\infty}^{+\infty} \mathrm{d}w\, \frac{\partial p\,(w,t)}{\partial t}\, e^{\imath s\, w} \;=\; \frac{1}{2}\int_{-\infty}^{+\infty} \mathrm{d}w\, \frac{\partial^2 p\,(w,t)}{\partial w^2}\, e^{\imath s\, w}\;.$$

First we derive a differential equation for the characteristic function by applying integration by parts twice.[25] The first and second partial integration steps yield

---

[24] We distinguish the two formally identical equations (3.55) and (3.56), because the interpretation is different: The first equation (3.55) describes the evolution of a probability distribution with the conservation relation $\int \mathrm{d}w\, p\,(w,t) = 1$, whereas the second equation (3.56) deals with a concentration profile, which fulfils $\int \mathrm{d}x\, c\,(x,t) = c_{\mathrm{tot}}$ corresponding to mass conservation. In case of the heat equation the conserved quantity is total heat.

[25] Integration by parts is a standard integration method in calculus. It is encapsulated in the formula

$$\int_a^b u(x)\, v'(x)\, \mathrm{d}x \;=\; u(x)\, v(x)\,\Big|_a^b \;-\; \int_a^b u'(x)\, v(x)\, \mathrm{d}x\;.$$

Characteristic functions are especially well suited for partial integration, because exponential functions, $v(x) = \exp(\imath s x)$, can be easily integrated and probability densities $u(x) = p\,(x,t)$ as well as their first derivatives $u(x) = \partial p\,(x,t)/\partial x$ vanish in the limits $x \to \pm\infty$.

$$\int_{-\infty}^{+\infty}\mathrm{d}w\,\frac{\partial p\,(w,t)}{\partial w}\,e^{\imath s\,w}=p\,(w,t)\,e^{\imath s w}\,\Big|_{-\infty}^{\infty}-\int_{-\infty}^{+\infty}\mathrm{d}w\,p\,(w,t)\,\frac{\partial e^{\imath s\,w}}{\partial w}=\imath s\,\phi(s,t)$$

$$\text{and}\quad\int_{-\infty}^{+\infty}\mathrm{d}w\,\frac{\partial^2 p\,(w,t)}{\partial w^2}\,e^{\imath s\,w}=-\,s^2\,\phi(s,t)\quad.$$

The function $p\,(w,t)$ is a probability density and accordingly has to vanish in the limits $w\to\pm\infty$. The same is true for the first derivatives, $\partial p\,(w,t)/\partial w$. Differentiation of $\phi(s,t)$ in equation (2.29) with respect to t and applying equation (3.55) we obtain

$$\frac{\partial\phi(s,t)}{\partial t}\;=\;-\frac{1}{2}\,s^2\,\phi(s,t)\tag{3.58}$$

Next we compute the characteristic function by integration and find:

$$\phi(s,t)\;=\;\phi(s,t_0)\cdot\exp\Big(-\frac{1}{2}\,s^2\,(t-t_0)\Big)\;.\tag{3.59}$$

Insertion of the initial condition $\phi(s,t_0)=\exp\,(\imath s\,w_0)$ completes the characteristic function

$$\phi(s,t)\;=\;\exp\Big(\imath s\,w_0-\frac{1}{2}\,s^2\,(t-t_0)\Big)\tag{3.60}$$

and eventually we find the probability density through inverse Fourier transformation

$$p\,(w,t)\;=\;\frac{1}{\sqrt{2\pi\,(t-t_0)}}\,\exp\left(-\frac{(w-w_0)^2}{2\,(t-t_0)}\right)\quad\text{with}\quad p\,(w,t_0)=\delta(w-w_0)\,.\tag{3.61}$$

The density function of the Wiener process is a normal distribution with expectation value and variance,

$$\mathrm{E}\Big(\mathcal{W}(t)\Big)\;=\;w_0\quad\text{and}\quad\mathrm{var}\Big(\mathcal{W}(t)\Big)\;=\;\mathrm{E}\Big(\big(\mathcal{W}(t)-w_0\big)^2\Big)\;=\;t-t_0\;,\tag{3.62}$$

respectively. The standard deviation, $\sigma(t)=\sqrt{t-t_0}$, is proportional to the square root of the time elapsed since the start of the process, $t-t_0$, and fulfils the famous $\sqrt{t}$-law. Starting the Wiener process at time $t_0=0$ at the origin $w_0=0$ yields $\mathrm{E}\big(\mathcal{W}(t)\big)=0$ and $\sigma\big(\mathcal{W}(t)\big)^2=t$. An initially sharp distribution spreads in time as illustrated in figure 3.8, and this is precisely what is experimentally observed in diffusion. The infinite time limit of (3.61) is a uniform distribution $\mathcal{U}(w)=0$ on the whole real axis and hence $p\,(w,t)$ vanishes in the limit $t\to\infty$,.

Although the expectation value $\mathrm{E}\big(\mathcal{W}(t)\big)=w_0$ is well defined and independent of time in the sense of a martingale, the mean square $\mathrm{E}\big(\mathcal{W}(t)^2\big)$ becomes infinite as $t\to\infty$. This implies that the individual trajectories, $\mathcal{W}(t)$, are extremely variable and diverge after short time (see, for example, the five trajectories of the forward equation in figure 3.3). We shall encounter

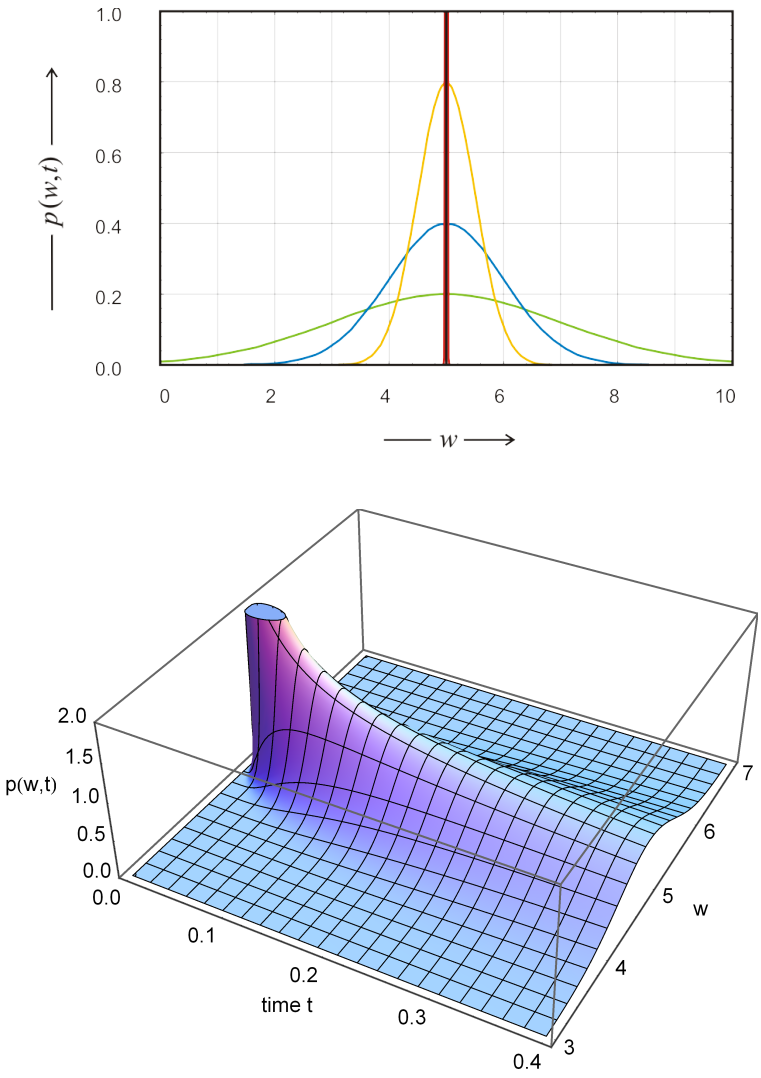


Fig. 3.8 Probability density of the Wiener process. In the figure we show the conditional probability density of the Wiener process, which is identical with the normal distribution (figure 1.21),

$$p\,(w,t) = p\,(w,t|w_0,t_0) = \exp\Big(-(w-w_0)^2/\big(2(t-t_0)\big)\Big)/\sqrt{2\pi(t-t_0)}.$$

The initially sharp distribution, $p\,(w,t_0|w_0,t_0) = \delta(w-w_0)$ spreads with increasing time until it becomes completely flat in the limit $t \to \infty$. The values used are $w_0 = 5\,[l.u.]$, $t_0 = 0$, and t =0 (black), 0.01 (red), 0.5 (yellow), 1.0 (blue), and 2.0 $[t.u.]$ (green). The lower picture shows a 3D plot of the density function.

such a situation with finite mean but diverging variance also in biology in the case of multiplication as a pure birth and death process (chapter 5). The expectation value although well defined looses its meaning in practice when the standard deviation becomes larger than the mean.

Continuity of sample paths of the Wiener process has been discussed already in subsection 3.2. Here we present proofs for two more features of the Wiener process: (i) individual trajectories, although being continuous, are nowhere differentiable and (ii) the increments of the Wiener Process are independent of each other. The nondifferentiability of the trajectories of the Wiener process has a consequence for the physical interpretation as Brownian motion: The moving particle has no defined velocity. Independence of increments is indispensable for the integration of stochastic differential equations (section 3.4).

In order to show nondifferentiability we consider the convergence behavior of the difference quotient

$$\lim_{h\to 0} \left| \frac{\mathcal{W}(t+h) - \mathcal{W}(t)}{h} \right| ,$$

where the random variable $\mathcal{W}$ has the conditional probability (3.61). Ludwig Arnold [16, p.48] illustrates the nondifferentiability in a heuristic way: The difference quotient $\big(\mathcal{W}(t+h) - \mathcal{W}(t)\big)/h$ follows the normal distribution $\mathcal{N}(0, 1/|h|)$, which diverges as $h \downarrow 0$ – the limit of a normal distribution with exploding variance is undefined – and hence for every bounded measurable set S we have

$$P\Big(\big(\mathcal{W}(t+h) - \mathcal{W}(t)\big)/h \in \mathrm{S}\Big) \;\to\; 0 \;\text{ as }\; h \downarrow 0 \,.$$

Accordingly, the difference quotient cannot converge with nonzero probability to a random variable with finite value. The convergence can behavior be made more precise by using the law of the iterated logarithm (2.60): For almost every sample function and arbitrary ϵ in the interval $0 < \epsilon < 1$ as $h \downarrow 0$

$$\frac{\mathcal{W}(t+h) - \mathcal{W}(t)}{h} \;\geq\; (1-\epsilon)\sqrt{\frac{2\ln(\ln\big(1/h\big)\big)}{h}} \;\text{ infinitely often}$$

and simultaneously

$$\frac{\mathcal{W}(t+h) - \mathcal{W}(t)}{h} \;\leq\; (-1+\epsilon)\sqrt{\frac{2\ln(\ln\big(1/h\big)\big)}{h}} \;\text{ infinitely often} \,.$$

Since expressions on the r.h.s. approach $\pm\infty$ as $h \downarrow 0$, the difference quotient $\big(\mathcal{W}(t+h) - \mathcal{W}(t)\big)/h$ has for every fixed t. with probability one, the extended real line $[-\infty, +\infty]$ as its limit set of cluster points.

Because of the general importance of the Wiener process it is essential to present a proof for the statistical independence of nonoverlapping increments of $\mathcal{W}(t)$ [157, pp. 67,68]. We are dealing with a Markov process and hence we can write the joint probability as a product of conditional probabilities (3.16'), where $t_n - t_{n-1}, \ldots, t_1 - t_0$ are subintervals of the time span $t_n \geq t \geq t_0$

$$p\,(w_n, t_n; w_{n-1}, t_{n-1}; \ldots; w_0, t_0) \;=\; \prod_{i=0}^{n-1} p\,(w_{i+1}, t_{i+1} |\, w_i, t_i)\, p\,(w_0, t_0)\;.$$

Next we introduce new variables that are consistent with this partition: $\big(\Delta w_i \equiv \mathcal{W}(t_i) - \mathcal{W}(t_{i-1}), \Delta t_i \equiv t_i - t_{i-1}\big)\,\forall\, i = 1, \ldots, n$. Since $\mathcal{W}(t)$ is a also Gaussian process (see section 3.2.2.3) the probability density of any partition is normally distributed and we express the conditional probabilities in terms of (3.61):

$$p\,(\Delta w_n, \Delta t_n; \Delta w_{n-1}, \Delta t_{n-1}; \ldots; w_0, t_0) \;=\; \prod_{i=i}^{n} \frac{\exp\left(-\frac{\Delta w_i^2}{2\Delta t_i}\right)}{\sqrt{2\pi \Delta t_i}}\, p\,(w_0, t_0)\;.$$

The joint probability distribution is factorized into distributions from individual intervals and provided the intervals don't overlap the increments Δw_i are stochastically independent random variables in the sense of section 1.6.4, and accordingly, they are independent of the initial condition $\mathcal{W}(t_0)$. The independence relation is readily cast in precise form

$$\mathcal{W}(t) \,-\, \mathcal{W}(s) \;\text{ is independent of }\; \{\mathcal{W}(\tau)\}_{\tau \leq s} \;\text{ for any }\; 0 \leq s \leq t\;, \quad (3.63)$$

which will be used in the forthcoming sections on stochastic differential equations (section 3.4).

Applying equation (3.62) to the probability distribution within a partition we find for an the interval $\Delta t_k = t_k - t_{k-1}$:

$$\mathrm{E}\big(\mathcal{W}(t_k) - \mathcal{W}(t_{k-1})\big) \;=\; \mathrm{E}(\Delta w_k) \;=\; w_{k-1} \;\text{ and }\; \mathrm{var}(\Delta w_k) \;=\; t_k \,-\, t_{k-1}\,,$$

It is now straightforward to calculate the autocorrelation function, which is defined by

$$\begin{aligned} \langle \mathcal{W}(t)\mathcal{W}(s) | (w_0, t_0) \rangle \;&=\; \mathrm{E}\big(\mathcal{W}(t)\mathcal{W}(s) | (w_0, t_0)\big) \;= \\ &=\; \iint \mathrm{d}w_t \mathrm{d}w_s\, w_t w_s\, p\,(w_t, t; w_s, s | w_0, t_0)\;. \end{aligned} \quad (3.64)$$

Substraction and addition of $\mathcal{W}(s)^2$ inside the expectation value yields

$$\mathrm{E}\big(\mathcal{W}(t)\mathcal{W}(s) | (w_0, t_0)\big) \;=\; \mathrm{E}\Big(\big(\mathcal{W}(t) - \mathcal{W}(s)\big)\mathcal{W}(s)\Big) \;+\; \mathrm{E}\big(\mathcal{W}(s)^2\big)\;,$$

where the first term vanishes because of independence of the increments and the second term follows from (3.62):

$$\mathrm{E}\big(\mathcal{W}(t)\mathcal{W}(s)|(w_0,t_0)\big) \;=\; \min\{t-t_0, s-t_0\} \;+\; w_0^2 \;, \qquad (3.65)$$

and simplifies to $\mathrm{E}\big(\mathcal{W}(t)\mathcal{W}(s)\big) = \min\{t,s\}$ for $w_0 = 0$ and $t_0 = 0$. This expectation value reproduces also the diagonal element, the variance var, since for $s = t$ we find $\mathrm{E}\big(\mathcal{W}(t)^2\big) = t$. In addition, several other useful relations can be derived from the autocorrelation relation. We summarize:

$$\begin{aligned}
&\mathrm{E}\big(\mathcal{W}(t) - \mathcal{W}(s)\big) \;=\; 0\,,\; \mathrm{E}\big(\mathcal{W}(t)^2\big) \;=\; t\,,\; \mathrm{E}\big(\mathcal{W}(t)\mathcal{W}(s)\big) \;=\; \min\{t,s\}\,,\\
&\mathrm{E}\Big(\big(\mathcal{W}(t) - \mathcal{W}(s)\big)^2\Big) \;=\; \mathrm{E}\big(\mathcal{W}(t)^2\big) \;-\; 2\mathrm{E}\big(\mathcal{W}(t)\mathcal{W}(s)\big) \;+\; \mathrm{E}\big(\mathcal{W}(t)^2\big) \;=\\
&\qquad\qquad =\; t \;-\, 2\min\{t,s\} \;+\; s \;=\; |t-s|\;,
\end{aligned}$$

and remark that these results are not independent of the *càdlàg* convention for stochastic processes.

The Wiener process has the property of self-similarity: Assume that $\mathcal{W}_1(t)$ is a Wiener process. Then, for every $\gamma > 0$,

$$\mathcal{W}_2(t) \;=\; \mathcal{W}_1(\gamma t) \;=\; \sqrt{\gamma}\,\mathcal{W}_1(t)$$

is also a Wiener process. Accordingly, we can change the scale at will and the process remains a Wiener process. The power of the scaling factor is called the Hurst factor H (see sections 3.2.4 and 3.2.5), and accordingly the Wiener process has $H = 1/2$.

Solution of the diffusion equation by means of Fourier transform. The Fourier transform is as a convenient tool for deriving solutions of the diffusion equation, because transformation of derivatives results in algebraic equations in Fourier space and consecutive inverse transformation yields the desired answer.[26] In addition, the Fourier transform provides otherwise hard to obtain insights into problems.

The Fourier transform of a general derivative yields through integration by parts:

$$\begin{aligned}
\mathcal{F}\left(\frac{\mathrm{d}p\,(x)}{\mathrm{dx}}\right) \;&=\; \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{d}x\,p\,(x)\,e^{-\dot{u}\,kx} \;=\\
&=\; \frac{1}{\sqrt{2\pi}}\,p\,(x)\,e^{-\dot{u}\,kx}\,\Big|_{-\infty}^{\infty} \;+\; \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{d}x\,\dot{u}\,k\,p\,(x)\,e^{-\dot{u}\,kx} \;=\\
&=\; \dot{u}\,k\,\tilde{p}\,(k)\;.
\end{aligned}$$

[26] Integral transformations, in particular the Fourier and the Laplace transform, are standard techniques for solving ODEs and PDEs. For details we refer to mathematics handbooks for scientist, for example [123, pp. 89-96] and [382, pp. 449-451, 681-686].

The first term of integration by parts vanishes because $\lim_{x\to\pm\infty} p\,(x) = 0$, otherwise the probability could not be normalized. Application of the Fourier transform to higher derivatives requires multiple application of integration by parts an yields:

$$\mathcal{F}\left(\frac{\mathrm{d}^n p\,(x)}{\mathrm{dx}^n}\right) \;=\; (\imath\, k)^n\, \tilde{p}\,(k)\;. \tag{3.66}$$

Since t is handled like a constant in the Fourier transformation and in the differentiation by x and the two linear operators $\mathcal{F}$ and $\mathrm{d}/\,\mathrm{dt}$ can be interchanged without changing the result we find for the Fourier transformed diffusion equation:

$$\frac{\mathrm{d}\tilde{p}\,(k,t)}{\mathrm{dt}} \;=\; -\,D\,k^2\,\tilde{p}\,(k,t)\;. \tag{3.67}$$

The original PDE has become an ODE, which can be readily solved now and which yields

$$\tilde{p}\;(k,t) \;=\; \tilde{p}\,(k,0)\,\sqrt{\frac{Dt}{\pi}}\,e^{-D\,k^2\,t}\;. \tag{3.68}$$

This equation corresponds to a relaxation process with a relaxation time $\tau_{\mathrm{R}} = D\,k^2$ where k is the wave number with dimension $[l^{-1} = cm^{-1}]$.[27] The solution of the diffusion diffusion equation is then obtained by inverse Fourier transformation

$$p\,(x,t) \;=\; \frac{1}{\sqrt{4\pi D\,t}}\;e^{-x^2/(4Dt)}\;. \tag{3.69}$$

The solution, of course, is identical with the solution of the Wiener process in equation (3.61).

Multivariate Wiener Process. The Wiener process is readily extended to higher dimension. For the multivariate Wiener process, defined as

$$\vec{\mathcal{W}}(t) \;=\; \Big(\mathcal{W}_1(t),\ldots,\mathcal{W}_n(t)\Big) \tag{3.70}$$

satisfying the Fokker-Planck equation

$$\frac{\partial p\,(\mathbf{w},t|\mathbf{w}_0,t_0)}{\partial t} \;=\; \frac{1}{2}\sum_i \frac{\partial^2}{\partial w_i^2}\,p\,(\mathbf{w},t|\mathbf{w}_0,t_0)\;. \tag{3.71}$$

The solution is a multivariate normal density

$$p\,(\mathbf{w},t|\mathbf{w}_0,t_0) \;=\; \frac{1}{\sqrt{2\pi\,(t-t_0)}}\,\exp\left(-\frac{(\mathbf{w}-\mathbf{w}_0)^2}{2\,(t-t_0)}\right)\;. \tag{3.72}$$

with mean $\mathrm{E}\big(\vec{\mathcal{W}}(t)\big) = \mathbf{w}_0$ and variance-covariance matrix

[27] For a system in 3D space the wave vector in reciprocal space is denoted by $\mathbf{k}$ and its length $|\mathbf{k}| = k$ is called the wave number.

$$\big(\boldsymbol{\Sigma}\big)_{ij} \;=\; \mathrm{E}\Big(\big(\mathcal{W}_i(t)-w_{0i}\big)\big(\mathcal{W}_j(t)-w_{0j}\big)\Big) \;=\; (t-t_0)\,\delta_{ij}\;,$$

where all off-diagonal elements – the proper covariances – are zero. Hence, Wiener processes along different Cartesian coordinates are independent.

Before we consider the Gauss process as a generalization of the Wiener process it seems useful to summarize the most prominent features:

The Wiener process $W = \big(\mathcal{W}(t), t \geq 0\big)$ is characterized by ten properties and definitions:

(1) initial condition $\mathcal{W}(t_0) = \mathcal{W}(0) \equiv 0$,
(2) trajectories are continuous functions of $t \in [0, \infty[$,
(3) expectation value $\mathrm{E}\big(\mathcal{W}(t)\big) \equiv 0$,
(4) correlation function $\mathrm{E}\big(\mathcal{W}(t)\mathcal{W}(s)\big) = \min\{t, s\}$,
(5) Gaussian property implies that for any $(t_1, \ldots, t_n)$ the random vector $\big(\mathcal{W}(t_1), \ldots, \mathcal{W}(t_n)\big)$ is a Gaussian process, ,
(6) moments $\mathrm{E}\big(\mathcal{W}(t)^2\big) = t$, $\mathrm{E}\big(\mathcal{W}(t) - \mathcal{W}(s)\big) = 0$, and $\mathrm{E}\Big(\big(\mathcal{W}(t) - \mathcal{W}(s)\big)^2\Big) = |t - s|$,
(7) increments of the Wiener process on non-overlapping intervals are independent, for $(s_1, t_1) \cap (s_2, t_2) = \emptyset$ the random variables $\mathcal{W}(t_2) - \mathcal{W}(s_2)$ and $\mathcal{W}(t_1) - \mathcal{W}(s_1)$ are independent,
(8) nondifferentiability of trajectories $\mathcal{W}(t)$,
(9) self-similarity of the Wiener process $\mathcal{W}_2(t) = \mathcal{W}_1(\gamma t) = \sqrt{\gamma}\mathcal{W}_1(t)$, and
(10) martingale property, for $\mathcal{W}_0^s = \mathcal{W}(u)\,\forall\, 0 \leq u \leq s\}$ we have $\mathrm{E}\big(\mathcal{W}(t)|\mathcal{W}_0^s\big) = \mathcal{W}(s)$ and $\mathrm{E}\Big(\big(\mathcal{W}(t) - \mathcal{W}(s)\big)^2 \;\big|\; \mathcal{W}_0^s\Big) = t - s$.

Out of these ten properties three will be most important for the goals we will pursue here: (i) continuity of sample paths, (ii) nondifferentiability of sample paths, and (iii) independence of increments.

Gaussian and AR(n)[28] *processes.* A generalization of Wiener processes is the Gaussian process $\mathcal{X}(\mathbf{t})$ with $\mathbf{t} \in \mathcal{T}$ where $\mathcal{T}$ may be a finite index set $\mathcal{T} = \{t_1, \ldots, t_n\}$ or the entire space of real numbers $\mathcal{T} = \mathbb{R}^d$ for continuous time. The integer d is the dimension of a problem, for example the number of inputs. The condition of a Gaussian process is that any finite linear combination of samples has a joint normal distribution: $(\mathcal{X}_\mathbf{t}, \mathbf{t} \in \mathcal{T})$ is Gaussian if and only if for every finite index set $\mathbf{t} = (t_1, \ldots, t_n)$ there exist real numbers μ_k and σ_{kl}^2 with $\sigma_{kk}^2 > 0$ such that

[28] An autoregressive process of order n is denoted by AR(n). The order n implies that n values of the stochastic variables at previous times are required to calculate the current value. An extension of the autoregressive model is the *autoregressive moving average* (ARMA) model.

$$\mathrm{E}\left(\exp\Big(\dot{u}\sum_{i=1}^{n} t_i \mathcal{X}_{t_i}\Big)\right) \;=\; \exp\Big(-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}^2\, t_i t_j + \dot{u}\sum_{i=1}^{n}\mu_i t_i\Big)\;, \qquad (3.73)$$

where μ_k $(i = 1, \ldots, n)$ are the mean value of the random variables $\mathcal{X}_i$ and $\sigma_{ij}^2 = \mathrm{cov}(\mathcal{X}_i, \mathcal{X}_j)$ with $i, j = 1, \ldots, n$ are the elements of the covariance matrix Σ. The Wiener process is a nonstationary special case of a Gaussian process since the variance grows with $\sqrt{t}$. The Ornstein-Uhlenbeck process to be discussed in the next section 3.2.2.3 is an example for a stationary Gaussian process. After an initial transient period it settles down to a process with time-independent mean, $\bar{\mu}$ and variance $\bar{\mu}_2$. In a nutshell, a Gaussian process can be characterized as a normal distribution migrating in state space and changing shape thereby.

According to Wold's decomposition named after Herman Wold [467] any stochastic process with stationary covariance can be expressed by a time series that is decomposed into an independent deterministic part and independent stochastic components

$$\mathcal{Y}_t \;=\; \eta_t \,+\, \sum_{j=0}^{\infty} b_j \mathcal{Z}_{t-j}\;, \qquad (3.74)$$

where η_t is a deterministic process, for example the solution of a difference equation, $\mathcal{Z}_{t-j}$ are independent and identically distributed (iid) random variables, and b_j are coefficients fulfilling $b_0 = 1$ and $\sum_{j=0}^{\infty} b_j^2 < \infty$. This representation is called the *moving average model*. A stationary Gaussian process, $\mathcal{X}_t$ with $t \in \mathcal{T} = \mathbb{N}$ can be written in the form of equation (3.74) with the condition that the variables $\mathcal{Z}$ are iid normal variables with mean $\bar{\mu} = 0$ and variance $\bar{\mu}_2 = \bar{\sigma}^2$, $\mathcal{Z}_{t-j} = \mathcal{W}_{t-j}$. Since the independent deterministic part can be easily removed nondeterministic or Gaussian linear processes

$$\mathcal{X}_t \;=\; \sum_{j=0}^{\infty} b_j \mathcal{W}_{t-j} \quad \text{with} \quad b_0 = 1 \qquad (3.75)$$

are frequently used in time series analysis. An alternative representation of times series called autoregression considers the stochastic process by making use of past values of the variable itself [189, 458]

$$\mathcal{X}_t \;=\; \varphi_1 \mathcal{X}_{t-1} \,+\, \varphi_2 \mathcal{X}_{t-2} \,+\, \ldots\, \varphi_n \mathcal{X}_{t-n} \,+\, \mathcal{W}_t\;. \qquad (3.76)$$

The process (3.76) is characterized as autoregressive of order n or as AR(n) process. Every AR(n) process has a linear representation of the kind shown in (3.75) where the coefficients b_j are obtained as functions of the φ_k-values [53]. In other words, for every Gaussian linear process there exists an AR(n) process such that the two autocovariance function can be made practically equal for all time differences $t_j - t_{j-1}$. For the first n time lags the match can be made perfect. An extension to continuous time is possible and special

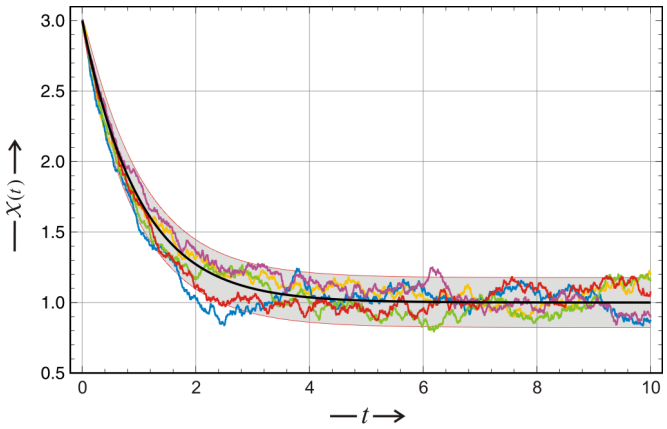


Fig. 3.9 The Ornstein-Uhlenbeck process. Individual trajectories of the process are simulated by $\mathcal{X}_{i+1} = \mathcal{X}_i\, e^{-k\,\vartheta} + \bar{\mu}(1-e^{-k\,\vartheta}) + \sigma\sqrt{\frac{1-e^{-2k\,\vartheta}}{2k}}\,(\mathcal{R}_{0,1}-0.5)$, where $\mathcal{R}_{0,1}$ is a random number drawn by a random number generator from the uniform distribution on the interval $[0,1]$. The figures shows several trajectories differing only in the choice of seeds for *Mersenne Twister* as random number generator. The lines represent the expectation value $\mathrm{E}\big(\mathcal{X}(t)\big)$ (black) and the functions $\mathrm{E}\big(\mathcal{X}(t)\big) \pm \sigma\big(\mathcal{X}(t)\big)$ (red). The area highlighted in gray is the confidence interval $E \pm \sigma$. Choice of parameters: $\mathcal{X}(0) = 3$, $\bar{\mu} = 1$, $k = 1$, $\sigma = 0.25$, $\vartheta = 0.002$ or total time $t_f = 10$. Seeds: 491 (yellow), 919 (blue), 023 (green), 877 (red), and 733 (violet). For the simulation of the Ornstein-Uhlenbeck model see [170, 436].

features of continuous time autoregressive models (CAR) are described, for example, in [54]. Finally we mention that AR(n) processes provide an excellent possibility to demonstrate the Markov property: An AR(1) process, $\mathcal{X}_t = \varphi\mathcal{X}_{t-1} + \mathcal{W}_t$ is Markovian in first order since the knowledge of $\mathcal{X}_{t-1}$ is sufficient to compute $\mathcal{X}_t$ and all further development in the future.

3.2.2.3 Ornstein-Uhlenbeck process

The Ornstein-Uhlenbeck process is named after two Dutch physicists Leonard Ornstein and George Uhlenbeck [433] and represents presumably the simplest stochastic process that approaches a stationary state with a defined variance.[29] The Ornstein-Uhlenbeck process found wide-spread applications, for example in economics for modeling irregular behavior of financial markets [443]. In physics it is among other applications a model for the velocity of a Brownian particle under the influence of friction. In essence, the Ornstein-

[29] The variance of the Wiener process diverges in the limit, $\lim_{t\to\infty} \mathrm{var}\big(\mathcal{W}(t)\big) = \infty$. The same is true for the Poisson process and the random walk, which are discussed in the next two sections.

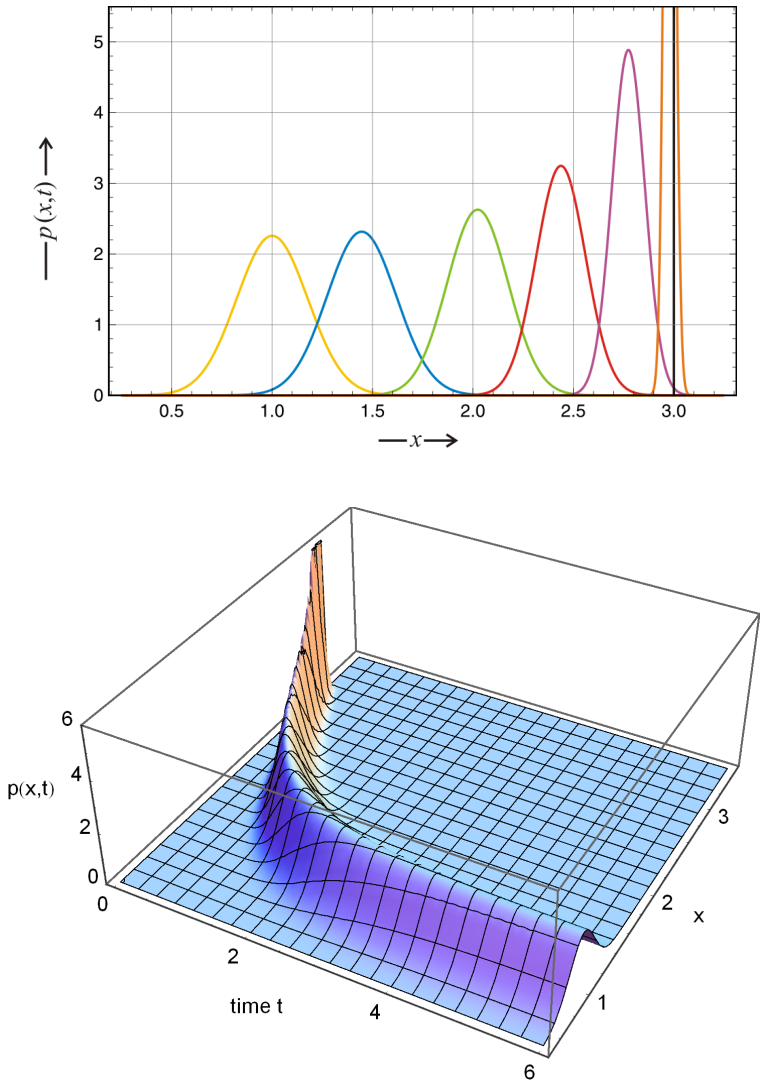


Fig. 3.10 The probability density of the Ornstein-Uhlenbeck process. Starting from the initial condition $p\,(x,t_0) = \delta(x-x_0)$ (black) the probability density (3.78) broadens and migrates until it reaches the stationary distribution (yellow). Choice of parameters: $x_0 = 3$, $\bar{\mu} = 1$, $k = 1$, and $\sigma = 0.25$. Times: $t = 0$ (black), 0.12 (orange), 0.33 (violet), 0.67 (green), 1.5 (blue), and 8 (yellow). The lower plot presents an illustration in 3D.

Uhlenbeck process describes exponential relaxation to a stationary state or to an equilibrium superimposed by a Wiener process. Figure 3.9 presents several trajectories of the Ornstein-Uhlenbeck process, which show nicely the drift and the diffusion component of the individual runs.

Fokker-Planck equation and solution of the Ornstein-Uhlenbeck process. The one-dimensional Fokker-Planck equation of the Ornstein-Uhlenbeck process for the probability density $p\,(x,t)$ of the random variable $\mathcal{X}(t)$ with the initial condition $p\,(x,t_0) = \delta(x - x_0)$ is of the form

$$\frac{\partial p\,(x,t)}{\partial t} \;=\; k\,\frac{\partial}{\partial x}\Big((x-\bar{\mu})\,p\,(x,t)\Big) \;+\; \frac{\sigma^2}{2}\,\frac{\partial^2 p\,(x,t)}{\partial x^2}\;, \tag{3.77}$$

with k is the rate parameter of the exponential decay, $\bar{\mu}$ the expectation value of the random variable in the long-time or stationary limit, $\bar{\mu} = \lim_{t\to\infty} \mathrm{E}\big(\mathcal{X}(t)\big)$, and $\bar{\mu}_2 = \lim_{t\to\infty} \mathrm{var}\big(\mathcal{X}(t)\big) = \sigma^2/(2k)$ being the stationary variance. For the initial condition $p\,(x,0) = \delta(x - x_0)$ the probability density can be obtained by standard techniques

$$p\,(x,t) \;=\; \sqrt{\frac{k}{\pi\sigma^2\,(1-e^{-2k\,t})}}\;\exp\left(-\frac{k}{\sigma^2}\,\frac{(x-\bar{\mu}-(x_0-\mu)e^{-kt})^2}{1-e^{-2kt}}\right)\;. \tag{3.78}$$

This expression can be easily checked by performing the two limits $t \to 0$ and $t \to \infty$. The first limit has to yield the initial conditions and it is indeed recalling a common definition of the Dirac delta-function.

$$\delta_\alpha(x) \;=\; \lim_{\alpha\to 0}\,\frac{1}{\alpha\sqrt{\pi}}\;e^{-x^2/\alpha^2}\;, \tag{3.79}$$

Inserting $\alpha^2 = \sigma^2(1-e^{-2kt})/k$ leads to

$$\lim_{t\to 0} p\,(x,t) \;=\; \delta(x-x_0)\;.$$

The long time limit of the probability density is calculated straightforwardly:

$$\lim_{t\to\infty} p\,(x,t) \;=\; \bar{p}\,(x) \;=\; \sqrt{\frac{k}{\pi\sigma^2}}\;e^{-k\,(x-\bar{\mu})^2/\sigma^2}\;, \tag{3.80}$$

which is a normal density with expectation value $\bar{\mu}$ and variance $\sigma^2/(2k)$. □
The evolution of probability density $p\,(x,t)$ from the δ-function at $t = 0$ to the stationary density $\lim_{t\to\infty} p\,(x,t)$ is shown in figure 3.10. The Ornstein-Uhlenbeck process is a stationary Gaussian process and has a representation as a first-order autoregressive $\big(AR(1)\big)$ process and this implies that it fulfils the Markov condition. It is illustrative to compare the three 3Dplots in figures 3.7, 3.8, and 3.10: The probability density of the Liouville process migrates according to the drift term, $\mu(t)$, but does not change shape that is the variance remains constant, the Wiener density stays in state space, $\mu = 0$,

but changes shape as the variance increases $\mu_2^2 = \sigma(t)^2 = t - t_0$, and finally the density of the Ornstein-Uhlenbeck process drifts and changes shape.

The Ornstein-Uhlenbeck process can be modeled efficiently also by a stochastic differential equation (SDE, see section 3.4.5):

$$\mathrm{d}x(t) \;=\; k\left(\bar{\mu} - x(t)\right)\,\mathrm{dt} \;+\; \sigma\, dW(t)\;. \tag{3.81}$$

The individual trajectories shown in figure 3.9 [170, 436] were simulated by means of the following equation

$$\mathcal{X}_{i+1} \;=\; \mathcal{X}_i\, e^{-k\,\vartheta} \;+\; \bar{\mu}(1 - e^{-k\,\vartheta}) \;+\; \sigma\sqrt{\frac{1 - e^{-2k\,\vartheta}}{2k}}\,(\mathcal{R}_{0,1} - 0.5)\;,$$

where $\vartheta = \Delta t/n_{st}$ is the number of steps per time interval Δt. The probability density can be derived, for example, from a sufficiently large ensemble of simulated trajectories. Expectation value and variance of the random variable $\mathcal{X}(t)$ can be calculated directly from the solution of the SDE (3.81) as shown in section 3.4.5.

Stationary solutions of Fokker-Planck equations. Often one is mainly interested in the long-time solution of a stochastic process and then the stationary solution of a Fokker-Planck equation, provided it exists, may be calculated directly. At stationarity time independence of the two functions $A(x,t)$ and $B(x,t)$ is assumed. We shall be dealing here with the one-dimensional case and consider the Ornstein-Uhlenbeck process as an example. We start by setting the time derivative of the probability density equal zero:

$$\begin{aligned}
\frac{\partial p\,(x,t)}{\partial t} \;&=\; 0 \;=\; -\frac{\partial}{\partial x}A(x)\,p\,(x,t) \;+\; \frac{1}{2}\frac{\partial^2}{\partial x^2}B(x)\,p\,(x,t) \quad \text{yielding} \\
A(x)\,\bar{p}\,(x) \;&=\; \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}x}\big(B(x)\,\bar{p}\,(x)\big)\;.
\end{aligned}$$

A convenient to integrate expression is obtained by means of a little trick [383, p. 98]:

$$\begin{aligned}
A(x)\,\bar{p}\,(x) \;&=\; \frac{A(x)}{B(x)}\left(B(x)\,\bar{p}\,(x)\right) \;=\; \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(B(x)\,\bar{p}\,(x,t)\big)\,, \\
\frac{\mathrm{d}\ln\big(B(x)\,\bar{p}\,(x)\big)}{\mathrm{d}x} \;&=\; \frac{2A(x)}{B(x)}\,, \quad \text{and} \quad B(x)\,\bar{p}\,(x) \;=\; \exp\!\left(2\int_0^x \frac{A(\xi)}{B(\xi)}\,\mathrm{d}\xi\right)\cdot\,\gamma\;,
\end{aligned}$$

where the factor γ arises from integration constants. Finally, we obtain

$$\bar{p}\,(x) \;=\; \frac{N}{B(x)}\,\exp\!\left(2\int_0^x \frac{A(\xi)}{B(\xi)}\,\mathrm{d}\xi\right) \tag{3.82}$$

with the integration constant absorbed into the normalization factor N, which takes care that the probability conservation relation $\int_{-\infty}^{\infty} \bar{p}\,(x) = 1$ is fulfilled. As a rule the calculation of N is straightforward with specific examples.

As an illustrative case we calculate the stationary probability density for the Ornstein-Uhlenbeck process. For $A(x) = -k(x - \bar{\mu})$ and $B(x) = \sigma^2$ we find

$$\bar{p}\,(x) \;=\; \frac{N}{\sigma^2}\, e^{-k(x-\bar{\mu})^2/\sigma^2} \quad \text{and} \quad N \;=\; 1 \Big/ \left(\int_{-\infty}^{\infty} e^{-k(x-\bar{\mu})^2/\sigma^2} \mathrm{d}x \;/\; \sigma^2 \right) \;,$$

yields the final result that reproduces the previous result obtained from the time dependent density by taking the limit $t \to \infty$

$$\bar{p}\,(x) \;=\; \sqrt{\frac{k}{\pi\sigma^2}}\, e^{-k\,(x-\bar{\mu})^2/\sigma^2} \;. \tag{3.80'}$$

We emphasize once more that this result was obtained without making use of the time dependent probability density $p\,(x,t)$ and the approach allows for the calculation of stationary solution also in cases where $p\,(x,t)$ has not been derived.

3.2.2.4 Poisson process

The three processes discussed so far in this section were all dealing with continuous variables and their probability densities. We continue by presenting two examples of processes dealing with discrete variables and pure jump processes according to equation (3.46c), which are modeled by master equations: the Poisson process and the discrete, one-dimensional random walk. We are stressing once more, that master equations and related techniques are tailored for analyzing and modelling stochasticity at low particle numbers, and are therefore of particular importance in biology and chemistry.

The master equation (3.46c) is rewritten for the discrete case by replacing the integral by a summation[30]

$$\begin{aligned} \frac{\partial p\,(x,t)}{\partial t} \;&=\; \int \mathrm{d}z \, \Big(\mathrm{W}(x\,|z,t)\, p\,(z,t) - \mathrm{W}(z\,|x,t)\, p\,(x,t) \Big) \;\Rightarrow \\ \frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;&=\; \sum_{m=0}^{\infty} \Big(\mathrm{W}(n\,|m,t)\, P_m(t) - \mathrm{W}(m\,|n,t)\, P_n(t) \Big) \;, \end{aligned} \tag{3.83}$$

[30] Riemann-Stieltjes integration converts the integral into a sum and since we are dealing with discrete events exclusively we use an index on the probability, $P_n(t)$, rather than writing the density with an additional variable, $P(n,t)$.

where we are assuming $n \in \mathbb{N}$, continuous time, $t \in \mathbb{R}_{\geq 0}$, and sharp initial conditions (n_0, t_0) or $P_n(t_0) = \delta_{n,n_0}$.[31] The matrix $W(m\,|n,t)$ is called the transition matrix that contains the probabilities attributed to jumps of variables, and from both equations follows that the diagonal elements, $W(n\,|n,t)$, cancel. The domain of the random variable is implicitly included in the range of integration or summation, respectively.

The Poisson process is commonly applied to model certain classes of independent cumulative random events. These may be, for example, electrons arriving at an anode, customers entering a shop, telephone calls arriving at a switch board or e-mails being registered on an account. Aside from independence, the requirement is an unstructured time profile of events or, in other words, the probability of occurrence of events is constant and does not depend on time. The cumulative number of these events is denoted by the random variable $\mathcal{N}(t) \in \mathbb{N}$. In other words $\mathcal{N}(t)$ is counting the number of arrivals and hence can only increase. The probability of arrival is assumed to be γ per unit time, or $\gamma\,\Delta t$ is the expected number of events recorded in a time interval of length Δt.

The Poisson process can also be seen as a *one-sided* random walk in the sense that the walker takes a step, for example to the right, with a probability γ within a unit time interval. Then, the elements of the transition matrix become

$$W(m|\,n,t) \;=\; \begin{cases} \gamma & \text{if} \;\; m \,=\, n+1\,, \\ 0 & \text{otherwise} \end{cases}\,, \tag{3.84}$$

where the probability that two or more arrivals occur within the differential time interval dt is of measure zero. In other words simultaneous arrivals of two or more events have zero probability. According to (3.46c') the master equation has the form

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; \gamma\Big(P_{n-1}(t) \,-\, P_n(t)\Big) \tag{3.85}$$

with the initial condition $P_n(t_0) = \delta_{n,n_0}$. In other words, the number of arrivals recorded before $t = t_0$ is n_0. The interpretation of (3.85) is straightforward: the increase in the probability to have n recorded events between time t and $t + \mathrm{dt}$ is proportional to the difference in probabilities between $n-1$ and n recorded events, because the elementary single arrival processes, $(n-1 \to n)$ and $(n \to n+1)$, increase or decrease the probability of n events, respectively.

---

[31] By δ_{ij} we denote the Kronecker delta named after the German mathematician Leopold Kronecker, which means

$$\delta_{ij} \begin{cases} 1 & \text{if} \;\; i = j \\ 0 & \text{if} \;\; i \neq j \end{cases}.$$

It represents the discrete analogue of the Dirac delta-function.

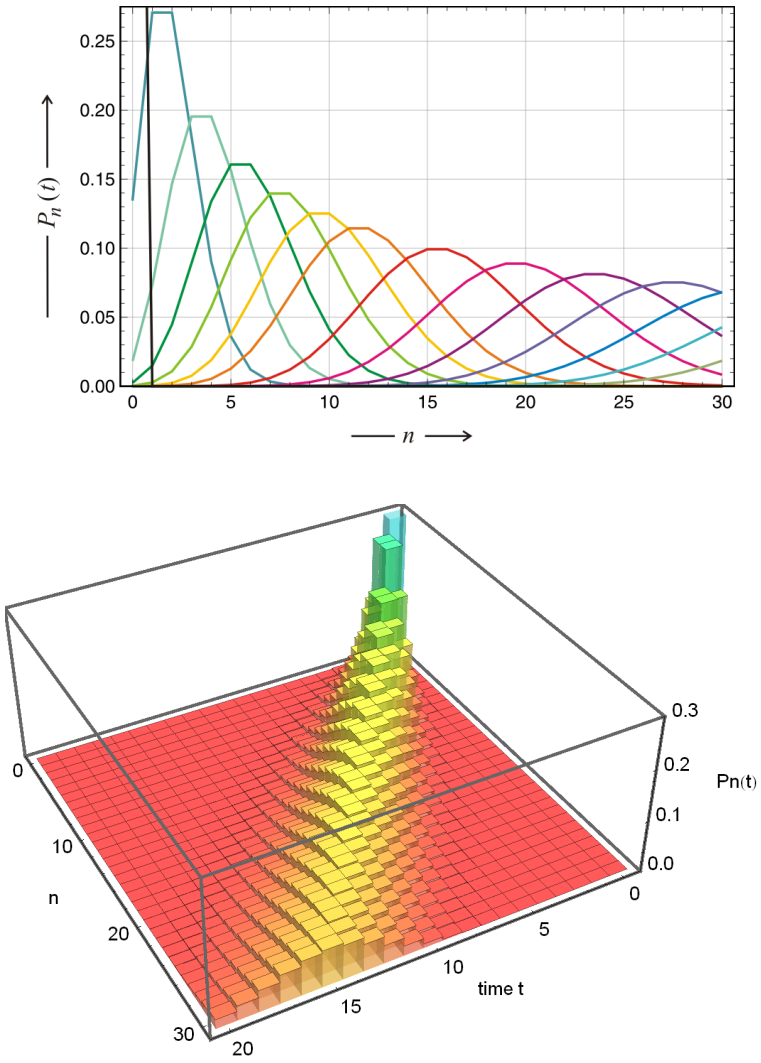


Fig. 3.11 Probability density of the Poisson process. The figures show the spreading of an initially sharp Poisson density, $P_n(t) = (\gamma t)^n e^{-\gamma t}/n!$, with time: $P_n(t) = p\,(n, t|n_0, t_0)$ with the initial condition $p\,(n, t_0|n_0, t_0) = \delta(n-n_0)$. In the limit $t \to \infty$ becomes the density completely flat. The values used are $\gamma = 2\,[t.u.]^{-1}$, $n_0 = 0$, $t_0 = 0$, and t =0 (black), 1 (sea green), 2 (mint green), 3 (green), 4 (chartreuse), 5 (yellow), 6 (orange), 8 (red), 10 (magenta), 12 (blue purple), 14 (electric blue), 16 (sky blue), 18 (turquoise), and 20 $[t.u.]$ (martian green). The lower picture shows a discrete 3D plot of the density function.

The method of probability generating functions (section 2.2.1) is now applied for deriving solutions of the master equation (3.85). The probability generating function for the Poisson process is

$$g(s,t) \;=\; \sum_{n=0}^{\infty} P_n(t)\, s^n \;,\; |s| \le 1 \;\; \text{with} \;\; g(s,t_0) \;=\; s^{n_0} \;. \qquad (2.24'))$$

The time derivative of the generation function is obtained by insertion of equation (3.85)

$$\frac{\partial g(s,t)}{\partial t} \;=\; \sum_{n=0}^{\infty} \frac{\partial P_n(t)}{\partial t}\, s^n \;=\; \gamma \sum_{n=0}^{\infty} \Big(P_{n-1}(t) - P_n(t) \Big)\, s^n \;,$$

the first sums is readily evaluated

$$\sum_{n=0}^{\infty} \frac{\partial P_{n-1}(t)}{\partial t}\, s^n \;=\; s \sum_{n=0}^{\infty} \frac{\partial P_{n-1}(t)}{\partial t}\, s^{n-1} \;=\; s\, g(s,t)$$

and the second sum is identical to the definition of the generating function. This yields the equation for the generating function

$$\frac{\partial g(s,t)}{\partial t} \;=\; \gamma\,(s-1)\, g(s,t) \;. \qquad (3.86)$$

Since the equation does not contain a derivative with respect to the dummy variable s we are dealing with a simple ODE and the solution by conventional calculus is straightforward:

$$\int_{\ln g(s,t_0)}^{\ln g(s,t)} \mathrm{d} \ln g(s,t) \;=\; \int_{t_0}^{t} \gamma (s-1)\, \mathrm{dt} \;,$$

which yields

$$g(s,t) \;=\; s^{n_0}\, e^{\gamma(s-1)(t-t_0)} \;\; \text{or} \;\; g(s,t) \;=\; e^{\gamma(s-1)t} \;\text{for}\; (n_0 = 0, t_0 = 0) \quad (3.87)$$

with $g(s,0) = s^{n_0}$. The assumption $(n_0 = 0, t_0 = 0)$ is meaningful, because it implies that counting arrivals starts at time $t = 0$, and the expressions become especially simple: $g(0,t) = \exp(-\gamma t)$ and $g(s,0) = 1$. The individual probabilities $P_n(t)$ are obtained through expansion of the exponential function and equating the coefficients for the powers of s:

$$\exp\big(\gamma(s-1)t\big) \;=\; \exp\big(\gamma\, st\big)\, e^{-\gamma t} \;\; \text{and}$$

$$\exp\big(\gamma\, st\big) \;=\; 1 \;+\; s\,\frac{\gamma t}{1!} \;+\; s^2\,\frac{(\gamma t)^2}{2!} \;+\; s^3\,\frac{(\gamma t)^3}{3!} \;+\; \ldots \;,$$

and eventually we are obtaining the solution

$$P_n(t) \;=\; e^{-\gamma t}\,\frac{(\gamma t)^n}{n!} \;=\; e^{-\lambda}\,\frac{\lambda^n}{n!}\;, \tag{3.88}$$

which is the well-known Poisson distribution (2.32) with the expectation value $\mathrm{E}\big(\mathcal{N}(t)\big) = \gamma t = \lambda$ and variance $\mathrm{var}\big(\mathcal{N}(t)\big) = \gamma t = \lambda$. Since the standard deviation is $\sigma\big(\mathcal{N}(t)\big) = \sqrt{\gamma t}$ and accordingly the Poisson process fulfils perfectly the $\sqrt{N}$ relation for fluctuations.

It is easily verified that expectation value and variance can be directly obtained from the generating function through differentiation (2.25):

$$\begin{aligned}
\mathrm{E}\big(\mathcal{N}(t)\big) \;&=\; \frac{\partial g(s,t)}{\partial s}\Big|_{s=1} \;=\; \gamma\, t\;, \\
\mathrm{var}\big(\mathcal{N}(t)\big) \;&=\; \frac{\partial g(s,t)}{\partial s}\Big|_{s=1} + \frac{\partial^2 g(s,t)}{\partial s^2}\Big|_{s=1} - \left(\frac{\partial g(s,t)}{\partial s}\Big|_{s=1}\right)^2 \;=\; \gamma\, t\;,
\end{aligned} \tag{3.89}$$

We remark that equation (3.85) can be solved also by using the characteristic function (section 2.2.3), which will be applied for the purpose of illustration in solving the master equation of the one-dimensional random walk (section 3.2.4).

The Poisson process can be viewed from a slightly different perspective by considering the *arrival times*[32] of the individual independent events as random variables $\mathcal{T}_1, \mathcal{T}_2, \cdots$, where the random counting variable takes on the values $\mathcal{N}(t) \geq 1$ for $t \geq \mathcal{T}_1$ and, in general $\mathcal{N}(t) \geq k$ for $t \geq \mathcal{T}_k$. All arrival times $\mathcal{T}_k$ with $k \in \mathbb{N}_{>0}$ are positive if we assume that the process started at time $t = 0$. The number of arrivals before some fixed time, ϑ, is smaller than k if and only if the waiting time until the k-th arrival is larger than ϑ. Accordingly the two events $\mathcal{T}_k < \vartheta$ and $n(\vartheta) < k$ are equivalent and their probabilities are the same

$$P(\mathcal{T}_k > \vartheta) \;=\; P\big(n(\vartheta) < k\big)\;.$$

Now we consider the time before the first arrival, which is trivially the time until the first event happens

$$P(\mathcal{T}_1 > \vartheta) \;=\; P\big(n(\vartheta) < 1\big) \;=\; P\big(n(\vartheta) = 0\big) \;=\; e^{-\vartheta/\tau_w}\,\frac{(\vartheta/\tau_w)^0}{0!} \;=\; e^{-t/\tau_w}\,,$$

where we used equation(3.88) for calculating the distribution of first-arrival times. It is straightforward to show that the same relation holds for all inter-arrival times $\Delta\mathcal{T}_k = \mathcal{T}_k - \mathcal{T}_{k-1}$, which after normalization follow an exponential density $\varrho(t,\tau_w) = e^{-t/\tau_w}/\tau_w$ with $\tau_w > 0$ and $\int_0^\infty \varrho(t,\tau_w)\,\mathrm{dt} = 1$, and thus for each index k we have

$$P(\Delta\mathcal{T}_k \leq t) \;=\; 1 - e^{-t/\tau_w} \quad \text{and thus} \quad P(\Delta\mathcal{T}_k > t) \;=\; e^{-t/\tau_w}\;,\; t \geq 0\;.$$

[32] In the literature the synonym *waiting time* or for arrival time is common.

Now we can identified the parameter of the Poisson distribution as the reciprocal *mean waiting time* for an event, τ_w^{-1}, with

$$\tau_w \;=\; \int_0^\infty \mathrm{dt}\, t\, \varrho(t,\tau_w) \;=\; \int_0^\infty \mathrm{dt}\; \frac{t}{\tau_w}\, e^{-\vartheta/\tau_w}\;.$$

We shall use the exponential density in the calculation of expected times for the occurrence of chemical reactions modeled as first arrival times $\mathcal{T}_1$. Independence of the individual events implies the validity of

$$\begin{aligned} P(\Delta\mathcal{T}_1 > t_1, \ldots, \Delta\mathcal{T}_n > t_n) \;&=\; P(\Delta\mathcal{T}_1 > t_1) \cdot \ldots \cdot P(\Delta\mathcal{T}_n > t_n) \;= \\ &=\; e^{-(t_1+\ldots+t_n)/\tau_w}\,, \end{aligned}$$

which determines the joint probability distribution of the *inter-arrival* times $\Delta\mathcal{T}_k$'s. The expectation value of the incremental arrival times, or times between consecutive arrivals, is simply given by $\mathrm{E}(\Delta\mathcal{T}_k) = \tau_w$. Clearly, the larger τ_w is, the longer will be the mean inter-arrival time, and thus $1/\tau_w$ can be addressed as the intensity of flow. Compared with the previous derivation we have $1/\tau_w \equiv \gamma$.

For $\mathcal{T}_0 = 0$ and $n \geq 1$ we can readily calculate the cumulative random variable, the arrival time of the the nth arrival

$$\mathcal{T}_n \;=\; \Delta\mathcal{T}_1 \,+\, \ldots\, \Delta\mathcal{T}_n \;=\; \sum_{k=1}^{n} \Delta\mathcal{T}_k\;.$$

The event $\mathcal{I} = (\mathcal{T}_n \leq t)$ implies that the nth arrival has occurred before time t. The connection between the arrival times and the cumulative number of arrivals, $\mathcal{N}(t)$, is easily performed and illustrates the usefulness of the dual point of view:

$$P(\mathcal{I}) \;=\; P(\mathcal{T}_n \leq t) \;=\; P(\mathcal{N}(t) \geq n)\;.$$

More precisely, $\mathcal{N}(t)$ is determined by the whole sequence $\Delta\mathcal{T}_k\,(k \geq 1)$, and depends on the elements ω of the sample space through the individual inter-arrival times $\Delta\mathcal{T}_k$. In fact, we can compute the number of arrivals exactly as the joint probability to have recorded $n-1$ arrivals at time t and to record one arrival in the interval $(t, t+\Delta t]$ [435, pp. 70-72]

$$P(t \leq \mathcal{T}_n \leq t+\Delta t) \;=\; P\big(\mathcal{N}(t) = n-1\big) \,\cdot\, P\big(\Delta\mathcal{N}(\Delta t) = 1\big)\;.$$

Since the two time intervals $[0,t]$ and $(t, t+\Delta t]$ do not overlap, the two events are independent and the joint probability can be factorized. For the first factor we use the probability of a Poissonian distribution and the second factor follows simply from the definition of the parameter γ:

$$P(t \leq \mathcal{T}_n \leq t+\Delta t) \;=\; \frac{e^{-\gamma t}\,(\gamma t)^{n-1}}{(n-1)!} \,\cdot\, \gamma\Delta t\;.$$

In the limit $\Delta t \to \mathrm{dt}$ we obtain the probability density of the nth arrival time

$$f_{\mathcal{T}_n}(t) \;=\; \frac{\gamma^n t^{n-1}}{(n-1)!}\, e^{-\gamma t} \;, \tag{3.90}$$

which is known as *Erlang* distribution named after the Danish mathematician Agner Karup Erlang. It is straightforward now to compute the expectation value of the nth waiting time:

$$\mathrm{E}(\mathcal{T}_n) \;=\; \int_0^\infty t\, \frac{\gamma^n t^{n-1}}{(n-1)!}\, e^{-\gamma t}\, \mathrm{dt} \;=\; \frac{n}{\gamma} \;, \tag{3.91}$$

which is another linear relation: The nth waiting time is proportional to n with the proportionality factor being the reciprocal rate parameter $1/\gamma$.

The Poisson process is characterized by three properties: (i) the observations occur one at a time, (ii) the numbers of observations in disjoint time intervals are independent random variables, and (iii) the distribution of $\mathcal{N}(t+\Delta t) - \mathcal{N}(t)$ is independent of t. Then there exists a constant $\lambda > 0$ such that for $\Delta t = t - \tau > 0$ the difference $\mathcal{N}(t) - \mathcal{N}(\tau)$ is Poisson distributed with the parameter $\lambda\, \Delta t$

$$P\Big(\mathcal{N}(t+\Delta t) = k\Big) \;=\; \frac{\big(\lambda\, \Delta t\big)^k}{k!}\, e^{-\lambda\, \Delta t} \;,$$

for $\lambda = 1$ the process $\mathcal{N}(t)$ is a *unit* or rate one Poisson process, and the expectation value is $\mathrm{E}\big(\mathcal{Y}(t)\big) = \Delta t$. In other words the mean number of events per unit time, $\Delta t = 1$ is one. It $\mathcal{Y}(t)$ is a unit Poisson process and $\mathcal{Y}_\lambda(t) \equiv \mathcal{Y}(\lambda t)$, then $\mathcal{Y}_\lambda$ is a Poisson process with parameter λ. A Poisson process is an example of a *counting process* $\mathcal{N}(t)$ with $t \geq 0$ that fulfils three properties: (i) $\mathcal{N}(t) \geq 0$, (ii) $\mathcal{N}(t) \in \mathbb{N}$, and (iii) if $\tau \leq t$ then $\mathcal{N}(\tau) \leq \mathcal{N}(t)$. The number of events occurring during the time interval $(\tau, t]$ with $\tau < t$ is $\mathcal{N}(t) - \mathcal{N}(\tau)$.

3.2.3 Master equations

Master equations are applied for modeling stochastic processes on discrete sample spaces, $\mathcal{N}(t) \in \mathbb{N}$, and we have been dealing already with one particularly example, the occurrence of independent in form of the Poisson process section 3.2.2.4. Because of their general importance in particular in chemical kinetics and population dynamics in biology we shall present here a more detailed discussion of the properties and different versions of master equations.

3.2.3.1 General master equations

The master equations we are considering here describe continuous time processes, $t \in \mathbb{R}$. Then, the starting point is the dCKE for pure jump processes (3.46c) with the integral converted into a sum by Riemann-Stieltjes integration (section 3.2.2.4)

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; \sum_{m=0}^{\infty}\Big(W(n\,|\,m,t)\,P_m(t) - W(m\,|\,n,t)\,P_n(t)\Big)\,;\;\; n,m \in \mathbb{N}\;, \quad (3.83)$$

where we have implicitly assumed sharp initial conditions: $P_n(t_0) = \delta_{n,n_0}$. The *transition probabilities* $W(n\,|\,m,t)$ form a – possibly infinite – transition matrix. In all realistic cases, however, we shall be dealing with a finite *state space*: $m,n \in \{0,1,\ldots,N\}$ – this is tantamount to saying we are always dealing with a finite numbers of molecules in chemistry or to stating that population sizes in biology are finite. Since the off-diagonal elements of the transition matrix represent probabilities they are nonnegative by definition: $\mathrm{W} \doteq \{W_{nm};\, n,m \in \mathbb{N}_{\geq 0}\}$ (figure 3.12). The diagonal elements $W(n\,|\,n,t)$ cancel in the master equation and hence can be defined at will without changing the dynamics of the process. Two definitions are in common use:
(i) normalization of matrix elements

$$\sum_m W_{mn} \;=\; 1 \;\;\text{and}\;\; W_{nn} \;=\; 1 - \sum_{m\neq n} W_{mn}\;, \quad (3.92a)$$

and accordingly W is a stochastic matrix. This definition is applied, for example, in the mutation selection problem [105], or
(ii) annihilation of diagonal elements

$$\sum_m W_{mn} \;=\; 0 \;\;\text{and}\;\; W_{nn} \;=\; -\sum_{m,m\neq n} W_{mn}\;, \quad (3.92b)$$

which is used, for example, in the compact from of the master equation (3.83') and in several applications among them phylogeny.

Transition probabilities in the general master equation (3.83) are assumed to be time dependent. Most frequently we shall, however, assume that they do not depend on time and use $W_{nm} = W(n|m)$. A Markov process in general and a master equation in particular are called *time homogeneous* if the transition matrix W does not depend on time.

Formal solution of the master equation. Insertion of the annihilation condition (3.92b) into equation (3.83) leads to a compact form of the master equation

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; \sum_m W_{nm}P_m(t)\;. \quad (3.83')$$

Introducing vector notation, $\mathbf{P}(t)^{\mathrm{t}} = \big(P_1(t),\ldots,P_n(t),\ldots\big)$, we obtain

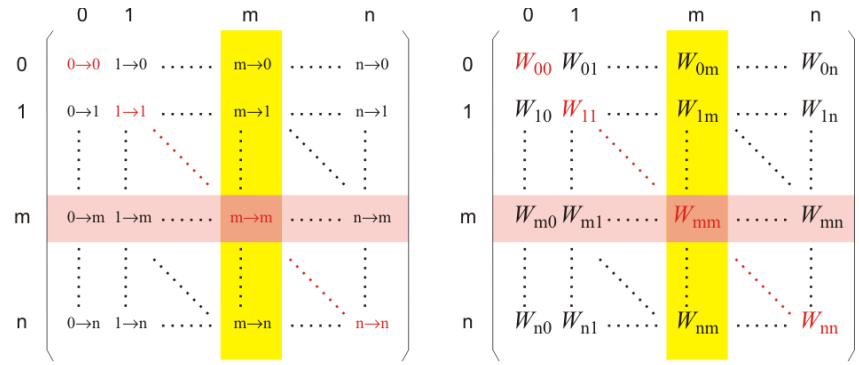


Fig. 3.12 The transition matrix of the master equation. The figure is intended to clarify meaning and handling of the elements of transition matrices in master equations. The matrix on the l.h.s. shows the individual transitions that are described by the corresponding elements of the transitions matrix $\mathrm{W} = \{W_{ij}; i, j = 0, 1, \ldots, n\}$. The elements in a given row (highlighted light red) contain all transitions going into one particular state 'm', and they are responsible for the differential change in probabilities: $\mathrm{d}P_m(t)/\,\mathrm{dt} = \sum_k W_{mk} P_k(t)$. The elements in a column (highlighted yellow) quantify all probability flows going out from state 'm' and their sums are involved in conservation of probabilities. The diagonal elements (red) cancel in master equations (3.83), hence they don't change probabilities, and need not be specified explicitly. For writing master equations in compact form (3.83') diagonal elements are defined by the annihilation convention, $\sum_k W_{km} = 0$. The summation of the elements in a column is also used in the definition of *jump moments*.

$$\frac{\mathrm{d}\mathbf{P}(t)}{\mathrm{dt}} \;=\; \mathrm{W} \times \mathbf{P}(t)\;. \tag{3.83"}$$

With the initial condition $P_n(0) = \delta_{n,n_0}$ stated above and a time independent transition matrix W we can solve equation (3.83") in formal terms for each n_0 by applying linear algebra and obtain

$$P(n, t|n_0, 0) \;=\; \Big(\exp\left(\mathrm{W}\,t\right)\Big)_{n,n_0}\;,$$

where the element (n, n_0) of the matrix $\exp(\mathrm{W}t)$ is the probability to have n particles at time t, $\mathcal{N}(t) = n$, when there were n_0 particles at time $t_0 = 0$. The computation of a matrix exponential is a quite elaborate task. If the matrix is diagonalizable, $\Lambda = \mathrm{T}^{-1} \cdot \mathrm{W} \cdot T$ with

$$\Lambda \;=\; \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}\;,$$

the exponential can be obtained by $e^{\mathrm{W}} = \mathrm{T}e^{\Lambda}\mathrm{T}^{-1}$. Apart from special cases the diagonalization of a matrix can be done analytically only in rather few low-dimensional cases and in general, one is restricted to numerical methods.

Jump moments. It is often convenient to express changes in particle numbers in terms of the so-called *jump moments* [343, 411, 439]

$$\alpha_p(n) \;=\; \sum_{m=0}^{\infty} (m-n)^p \, W(m|n)\,; \;\; p = 1, 2, \ldots \;. \tag{3.93}$$

The usefulness of the first two jump moments with $p = 1, 2$ is readily demonstrated: We multiply equation (3.83) by n and obtain through summation:

$$\begin{aligned}
\frac{\mathrm{d}\langle n\rangle}{\mathrm{dt}} \;&=\; \sum_{n=0}^{\infty} n \sum_{m=0}^{\infty} \Big(W(n|m)\, P_m(t) \;-\; W(m|n)\, P_n(t) \Big) \;= \\
&=\; \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} m\, W(m|n)\, P_n(t) \;-\; \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} n \, (W(m|n)\, P_n(t) \;= \\
&=\; \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} (m-n)\, (W(m|n)\, P_n(t) \;=\; \langle \alpha_1(n)\rangle \;.
\end{aligned}$$

Since the variance $\mathrm{var}(n) = \big\langle \big(n - \langle n\rangle\big)^2 \big\rangle$ involves $\langle n^2\rangle$, we need the time derivative of the second raw moment, $\widehat{\mu}_2 = \langle n^2\rangle$, and obtain it through multiplication of equation (3.93) with $p = 2$ by n^2 and summation

$$\begin{aligned}
\frac{\mathrm{d}\langle n^2\rangle}{\mathrm{dt}} \;&=\; \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} (m^2 - n^2)\, (W(m|n)\, P_n(t) \;= \\
&=\; \langle a_2(n)\rangle \;+\; 2\,\langle n\, a_1(n)\rangle \;.
\end{aligned}$$

Addition of the term $\mathrm{d}(\langle n\rangle^2)/\,\mathrm{dt} = 2\,\langle n\rangle\, \mathrm{d}\,\langle n\rangle/\,\mathrm{dt}$ yields the expression for the evolution of the variance and for the first two moments we finally obtain:

$$\frac{\mathrm{d}\langle n\rangle}{\mathrm{dt}} \;=\; \langle \alpha_1(n)\rangle \quad \text{and} \tag{3.94a}$$

$$\frac{\mathrm{d}\,\mathrm{var}(n)}{\mathrm{dt}} \;=\; \langle \alpha_2(n)\rangle \;+\; 2\,\Big(\langle n\, a_1(n)\rangle - \langle n\rangle\, \langle a_1(n)\rangle \Big) \;. \tag{3.94b}$$

Expression (3.94a) does not represent a closed equation for $\langle n\rangle$ since its solution involves higher moments of n. If $\alpha_1(n)$ is a linear function, however, the two summations, Σ_m for the jump moment and Σ_n for the expectation value, are interchangeable and after interchanging we obtain the simple ODE

$$\frac{\mathrm{d}\langle n\rangle}{\mathrm{dt}} \;= \alpha_1\big(\langle n\rangle\big) \;, \tag{3.94a'}$$

which can be integrated directly to yield the expectation value $\langle n(t)\rangle$ (see next section 3.2.3.2). Otherwise – in nonlinear systems – the expectation value does not coincide with the deterministic solution (see, for example, section 4.3), or in other words initial values of moments higher than the first one are required for computing the time course of the expectation value.

Nicholas van Kampen [439] provides also a straightforward approximation derived from an expansion series of $\alpha_1(n)$ in $\big(n - \langle n\rangle\big)$ with a break off after the second derivative

$$\frac{\mathrm{d}\langle n\rangle}{\mathrm{dt}} \;=\; \alpha_1\big(\langle n\rangle\big) \;+\; \frac{1}{2}\,\mathrm{var}(n)\,\frac{\mathrm{d}^2}{\mathrm{d}n^2}\,\alpha_1\big(\langle n\rangle\big)\;. \tag{3.94a"}$$

A similar and consistent approximation for the time dependence of the variance reads

$$\frac{\mathrm{d}\,\mathrm{var}(n)}{\mathrm{dt}} \;=\; \alpha_2\big(\langle n\rangle\big) \;+\; 2\,\mathrm{var}(n)\,\frac{\mathrm{d}}{\mathrm{d}n}\,\alpha_1\big(\langle n\rangle\big)\;. \tag{3.94b"}$$

The two expressions together provide a closed equation for the calculation of expectation value and variance, and they visualize directly the necessity knowing initial fluctuations for the computation of the time course of expectation values.

In the derivation of the dCKE and the master equation we made the realistic assumption that the limit of infinitesimal time steps, $\lim \Delta t \to 0$, excludes the simultaneous occurrence of two or more jumps. The general master equation (3.83), however, allows for jumps of all sizes $\Delta n = n - m$ and in most cases this introduces an unnecessary complication. In the next section we shall make use of a straightforward simplification in form of *death-and-birth* processes that restricts the size of jumps, reduces the numbers of terms in the master equation, and makes the expressions for the jump moments much easier to handle.

3.2.3.2 Birth-and-death master equations

The concept of birth-and-death processes has been created in biology (section 5.2.1) and is based on the assumption that constant and finite numbers of individuals are produced, 'born', or disappear, 'die', in single events. Accordingly the jump size is a matter of the application be it in physics, chemistry or biology, and the information on it has to come from empirical observations. To give examples, in chemical kinetics the jump size is determined by the stoichiometry of the process, and in population biology the jump size for birth is the litter size[33] and it is commonly one for natural death.

Here we shall consider jump size as a feature for mathematical characterization of stochastic processes. Jump size is a matter of handling single

[33] The litter size is defined as the mean number of offspring produced by an animal in s single birth.

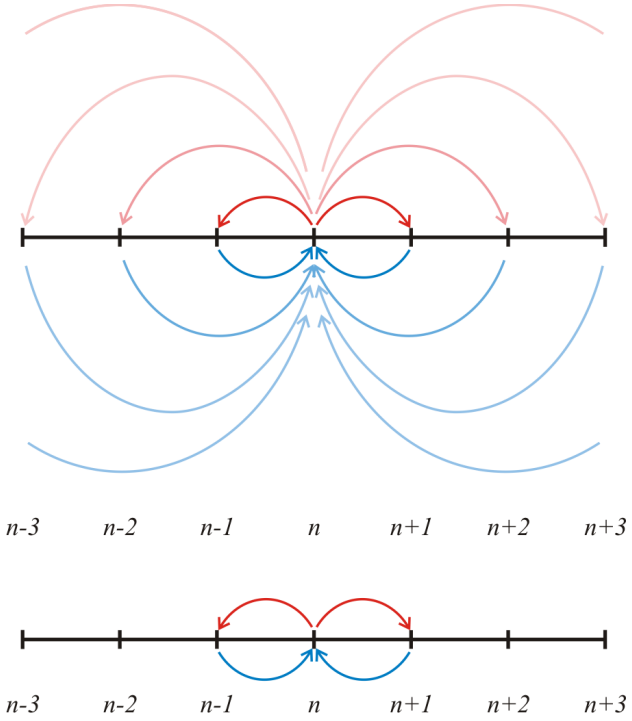


Fig. 3.13 Sketch of the transition probabilities in master equations. In the general master equation steps of any size are admitted (upper drawing) whereas in birth-and-death processes all jumps have the same size. The simplest and most common case is dealing with the condition that the particles *are born* and *die* one at a time (lower drawing), which is consistent with the derivation of the differential Chapman-Kolmogorov equation (section 3.2.1).

events, and we shall adopt the same procedure that we used in the derivation of the dCKE: We choose a sufficiently small time interval Δt for recording events such that the simultaneous occurrence of two events has a probability of measure zero. The resulting models are commonly called single step birth-and-death processes and the small time step Δt is addressed as *blind interval*, because the time resolution does not go beyond Δt. The difference in choosing steps between general and birth-and-death master equations is illustrated in figure 3.13 (see also section 4.6). In this chapter we shall restrict analysis and discussion to processes with a single variable and postpone the discussion of multivariate cases to chemical reaction networks handled in chapter 4.

Within the single step birth-and-death model the transition probabilities are reduced to neighboring states and we assume time independence

$$W(n|m) \;=\; W_{nm} \;=\; w_m^+\,\delta_{n,m+1} \,+\, w_m^-\,\delta_{n,m-1}\,, \quad \text{or}$$

$$W_{nm} \;=\; \begin{cases} w_m^+ & \text{if} \;\; m = n-1\,, \\ w_m^- & \text{if} \;\; m = n+1\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{3.95}$$

as we are dealing with only two allowed processes out of and into each state 'n' with the transition probabilities[34]

$$w_n^+ \;\; \text{for} \;\; n \;\rightarrow\; n+1 \;\; \text{and} \tag{3.96a}$$

$$w_n^- \;\; \text{for} \;\; n \;\rightarrow\; n-1 \;\; \text{respectively.} \tag{3.96b}$$

The notations *step-up* and *step-down* transitions for these two classes of events are self-explanatory. As a consequence of this simplification the transition matrix W becomes tridiagonal.

We have been dealing with birth-and-death processes already: In section 3.2.2.4 we discussed the Poisson process which can be understood as a birth-and-death process with zero death rate, or a birth process on $n \in \mathbb{N}_{\geq 0}$. The one-dimensional random walk (section 3.2.4) is a birth-and-death process with equal birth and death probabilities when the spatial coordinate is changed to a population variable and negative particle numbers are avoided. Modeling of chemical reactions by birth-and-death processes turns out to be a very useful approach for reaction mechanisms, which can be described by one step changes in a single variable.

The stochastic process can now be described by a birth-and-death master equation

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; w_{n-1}^+\,P_{n-1}(t) \,+\, w_{n+1}^-\,P_{n+1}(t) \,-\, (w_n^+ + w_n^-)\,P_n(t)\;. \tag{3.97}$$

There is no general technique that allows to find the time-dependent solutions of equation (3.97). Special cases, however, are important in chemistry and biology and therefore we shall present several examples later on. In section 5.2.1 we shall give also a detailed overview of the exactly solvable single step birth-and-death processes [176]. Nevertheless, it is possible to analyze the stationary case in full generality.

Provided a stationary solution of equation (3.97), $\lim_{t\to\infty} P_n(t) = \bar{P}_n$, exists, we can compute it in straightforward manner. We define a probability current $\varphi(n)$ for the n-th step in the series:

$$\begin{array}{ll} \text{Particle number} & 0 \rightleftharpoons 1 \rightleftharpoons \ldots \;\; \rightleftharpoons \;\; n-1 \rightleftharpoons n \;\; \rightleftharpoons \;\; n+1 \;\; \ldots \\ \text{Reaction step} & \qquad 1 \quad 2 \;\ldots\; n-1 \qquad\quad n \quad n+1 \qquad\quad \ldots\,, \end{array}$$

which is of the form

[34] Exceptions are the lowest and the highest state, $n = n_{\min}$ and $n = n_{max}$, which represent barriers for the process.

$$\varphi_n \;=\; w_n^- \, \bar{P}_n \;-\; w_{n-1}^+ \, \bar{P}_{n-1} \; . \tag{3.98}$$

Now, the conditions for the stationary solution are given by

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; 0 \;=\; \varphi_{n+1} \;-\; \varphi_n \;\; , \tag{3.99}$$

Restriction to positive particle numbers, $n \in \mathbb{N}_{\geq 0}$, implies $w_0^- = 0$ and $P_n(t) = 0$ for $n < 0$, which in turn leads to $\varphi_0 = 0$.

Now we add the vanishing flow terms according to equation (3.99) and obtain from the telescopic sum:

$$0 \;=\; \sum_{j=0}^{n-1} \varphi_{j+1} \;-\; \varphi_j \;\;=\;\; \varphi_n \;-\; \varphi_0 \; .$$

Thus we find $\varphi_n = 0$ for arbitrary n which leads to

$$\bar{P}_n \;=\; \frac{w_{n-1}^+}{w_n^-} \, \bar{P}_{n-1} \quad \text{and finally} \quad \bar{P}_n \;=\; \bar{P}_0 \prod_{j=1}^{n} \frac{w_{j-1}^+}{w_j^-} \; . \tag{3.100}$$

The vanishing flow condition $\varphi_n = 0$ for every reaction step at equilibrium is known in chemical kinetics as the principle of detailed balance, which has been formulated first by the American mathematical physicist Richard Tolman [430] (see also, for example, [157, pp.142-158]).

Calculating moments directly from master equations.[35] The simplification of the general master equation (3.83) by the restriction to single steps (3.97) provides a basis for the derivation of fairly simple expressions for the time derivatives of first and second moments. All calculations are facilitated by trivial but important equalities[36]

$$\sum_{n=-\infty}^{+\infty} (n-1) \, w_{n-1}^{\pm} \, P_{n-1}(t) = \sum_{n=-\infty}^{+\infty} n \, w_n^{\pm} \, P_n(t) = \sum_{n=-\infty}^{+\infty} (n+1) \, w_{n+1}^{\pm} \, P_{n+1}(t) \, ,$$

and we shall make use of these shifts in summation indices later in the search for solutions of master equations by means of probability generating functions. Through multiplication of $\mathrm{d}P_n/\,\mathrm{dt}$ with n, summation over n and using

[35] An excellent tutorial on this subject written by Bahram Houchmandzadeh is found in `http://www.houchmandzadeh.net/cours/Master_Eq/master.pdf`. Retrieved May 02, 2014.

[36] In general these equations hold also for summations from 0 to $+\infty$ if the corresponding physically meaningless probabilities are set equal zero by definition: $P_n(t) = 0 \, \forall \, n \in \mathbb{Z}_{<0}$.

$$\sum_{n=-\infty}^{\infty} (n+1)\, w_n^{\pm}\, P_n(t) \;=\; \sum_{n=-\infty}^{\infty} n\, w_n^{\pm}\, P_n(t) \;+\; \sum_{n=-\infty}^{\infty} w_n^{\pm}\, P_n(t)$$

we obtain for the expectation value:

$$\frac{\mathrm{d}\langle n\rangle}{\mathrm{dt}} \;=\; \sum_{n=-\infty}^{\infty} n\, \frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; \left\langle w_n^{+}\right\rangle - \left\langle w_n^{-}\right\rangle \;=\; \left\langle w_n^{+} - w_n^{-}\right\rangle \;. \qquad (3.101a)$$

The second raw moment, $\widehat{\mu}_2 = \left\langle n^2\right\rangle$, and the variance are derived by the analogous procedure – multiplication with n^2, summation, and substitution:

$$\begin{aligned}
\frac{\mathrm{d}\left\langle n^2\right\rangle}{\mathrm{dt}} \;&=\; \sum_{n=-\infty}^{\infty} n^2\, \frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; 2\,\left\langle n(w_n^{-} w_n^{-})\right\rangle + \left\langle w_n^{+} w_n^{-}\right\rangle \;, \\
\frac{\mathrm{d}\,\mathrm{var}(n)}{\mathrm{dt}} \;&=\; \frac{\mathrm{d}\left(\left\langle n^2\right\rangle - \langle n\rangle^2\right)}{\mathrm{dt}} \;=\; \frac{\mathrm{d}\left\langle n^2\right\rangle}{\mathrm{dt}} - \frac{\mathrm{d}\left(\langle n\rangle^2\right)}{\mathrm{dt}} \;= \\
&=\; 2\,\left\langle \big(n - \langle n\rangle\big)(w_n^{-} w_n^{-})\right\rangle + \left\langle w_n^{+} w_n^{-}\right\rangle \;, \qquad (3.101b)
\end{aligned}$$

Jump moments. Jump moments are substantially simplified by the assumption of single birth-and-death events:

$$\alpha_p\,(n) \;=\; \sum_{n=0}^{\infty} (m-n)^p\; W_{mn} \;=\; (-1)^p\, w_n^{-} \;+\; w_n^{+} \;,$$

which by neglecting fluctuations yields a rate equation for deterministic $\widehat{n}$

$$\frac{\mathrm{d}\widehat{n}}{\mathrm{dt}} \;=\; w_{\widehat{n}}^{+} \;-\; w_{\widehat{n}}^{-} \quad \text{with} \quad w_{\widehat{n}}^{\pm} = w_{\langle n\rangle}^{\pm} = \sum_{n=0}^{\infty} w_n^{\pm} P_n(t) \;, \qquad (3.102a)$$

as well as the two coupled simplified equations for the first two moments from equations (3.94a") and (3.94b"):

$$\frac{\mathrm{d}\langle n\rangle}{\mathrm{dt}} \;=\; w_{\langle n\rangle}^{+} \;-\; w_{\langle n\rangle}^{-} + \frac{1}{2}\,\mathrm{var}(n)\,\frac{\mathrm{d}^2}{\mathrm{d}n^2}\big(w_{\langle n\rangle}^{+} \;-\; w_{\langle n\rangle}^{-}\big) \quad \text{and} \qquad (3.102b)$$

$$\frac{\mathrm{d}\,\mathrm{var}(n)}{\mathrm{dt}} \;=\; w_{\langle n\rangle}^{+} \;+\; w_{\langle n\rangle}^{-} + 2\,\mathrm{var}(n)\,\frac{\mathrm{d}}{\mathrm{d}n}\big(w_{\langle n\rangle}^{+} \;-\; w_{\langle n\rangle}^{-}\big) \;. \qquad (3.102c)$$

It is now straightforward to show by example how linear jump moments simplify the expressions. For a linear birth-an-death process we find for step-up, step-down transitions, and jump moments

$$w_n^{+} \;=\; \lambda\, n \,,\; w_n^{+} \;=\; \lambda\, n \,, \quad \text{and} \quad \alpha_p\,(n) \;=\; \big(\lambda \;+\; (-1)^p\,\mu\big)\, n \;.$$

Twofold differentiation of the jump moments yields zero, the differential equations (3.102a) and (3.102b) are identical, and the solution is of the form

$$\langle n(t) \rangle \;=\; \widehat{n}(t) \;=\; \widehat{n}(0)\, e^{(\lambda-\mu)t} \;.$$

The expectation value of the stochastic variable $\langle n \rangle$ coincides with the deterministic variable $\widehat{n}$. We stress again that this coincidence requires linear step-up and step-down transition probabilities (see also section 4.2.1). More details on the linear birth-and-death process are found in section 5.2.1.

3.2.4 Continuous time random walks

The term *random walk* goes back to Karl Pearson [363] and is generally used for stochastic processes describing a walk in physical space with random increments. We have used to concept of a random walk in one dimension several times already in order to illustrate specific properties of stochastic processes (see, for example, section 3.1.1 and 3.1.3.2). Here we focus on the random walk itself and its infinitesimal step size limit, the Wiener process, which serves in physics at the same time as the basis of diffusion processes and as the model for white noise. For the sake of simplicity and access by analytical methods we shall be dealing here predominantly with the 1D random walk, although 2D and 3D walks are of similar or even greater importance in physics and chemistry.

In one and two dimensions the random walk is *recurrent* and this implies that each sufficiently long trajectory will visit every point in phase space, and it does this infinitely often if the trajectory is of infinite length. In particular, every trajectory will return to its origin. In three and higher dimensions this is not the case and the process is called *transient* therefore. A 3Dtrajectory revisits the origin in 34 % of the cases only, and this value decreases further in higher dimensions. Somewhat humoristically one can say a drunken sailor finds his way back home for sure, but a drunken pilot only in roughly one out of three trials. Random walks in one two and three dimensions are compared in figure fig:random-walks.

Random walk in one dimension. The 1D random walk is a classical problem of probability theory and science: A walker moves along a line and takes steps to the left or to the right with equal probability and length l, and regularly after a constant waiting time τ. The location of the walker is thus $n \cdot l$ with n being an integer, $n \in \mathbb{Z}$. In this form based on discrete space n and discrete time intervals τ we have used the random walk for illustrating the properties of a martingale in section 3.1.3.2. On the other hand we have studied already the continuous time and continuous space case as well in form of the Wiener process in section 3.2.2.2. Here we shall study now the *continuous time random walk* (CTRW) by keeping the step size discrete but assuming time to be continuous. In particular, a probability that the walker takes a step is defined, and the random walk is modeled by a master equation. In the forthcoming section 3.2.4 we shall consider random walks with probability distributions for the moves in space and time, step sizes and waiting times, respectively.

For the master equation we require transition probabilities per unit time, which are simply defined to be a constant, ϑ, for single steps and zero otherwise:

$$W(m|\,n,t) \;=\; \begin{cases} \vartheta & \text{if } \; m \,=\, n+1\,, \\ \vartheta & \text{if } \; m \,=\, n-1\,, \\ 0 & \text{otherwise} \end{cases} \;. \tag{3.103}$$

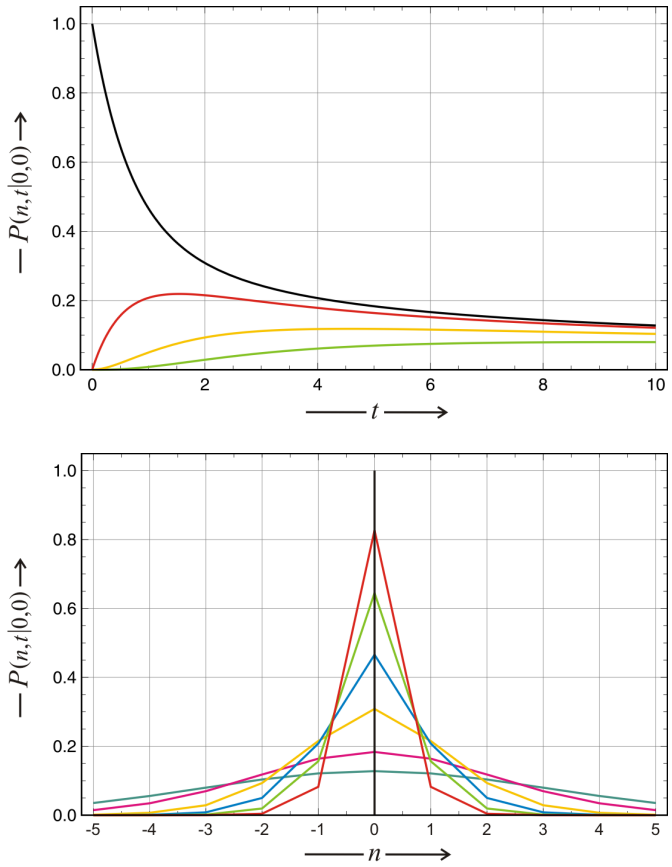


Fig. 3.14 Probability distribution of the random walk. The figure presents the conditional probabilities $P_n(t)$ of a random walker to be in location $n \in \mathbb{Z}$ at time t for the initial condition to be at $n = 0$ at time $t = t_0 = 0$. The upper part shows the dependence on t for given values of n: $n = 0$ (black), $n = 1$ (red), $n = 2$ (yellow), and $n = 3$ (green). The lower plot shows the probability distribution as a function of n at a given time t_k. Parameter choice: $\vartheta = 0.5$; $t_k = 0$ (black), 0.2 (red), 0.5 (green), 1 (blue), 2 (yellow), 5 (magenta), and 10 (cyan).

The master equation falls into the birth-and-death class and describes the evolution of the probability that the walker is at location $n \cdot l$ at time t,

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; \vartheta \left(P_{n+1}(t) \;+\; P_{n-1}(t) \;-\; 2\, P_n(t) \right) , \tag{3.104}$$

provided he started at location $n_0 \cdot l$ at time t_0: $P_n(t_0) = \delta_{n,n_0}$.

The master equation (3.104) can be solved by means of the time dependent characteristic function (see equations (2.29) and (2.29')):

$$\phi(s,t) \;=\; \mathrm{E}(e^{\dot{\imath}s\,n(t)}) \;=\; \sum_{n=-\infty}^{\infty} P_n(t)\,\exp\left(\dot{\imath}s\,n\right)\,. \tag{3.105}$$

Combining (3.104) and (3.105) yields

$$\frac{\partial\phi(s,t)}{\partial t} \;=\; \vartheta\left(e^{\dot{\imath}s}+e^{-\dot{\imath}s}-2\right)\phi(s,t) \;=\; 2\,\vartheta\left(\cosh(\dot{\imath}\,s)-1\right)\phi(s,t)\,.$$

Accordingly, the solution for the initial condition $n_0 = 0$ at $t_0 = 0$ is

$$\begin{aligned}\phi(s,t) \;&=\; \phi(s,0)\,\exp\Big(2\vartheta\,t\,\big(\cosh(\dot{\imath}\,s)-1\big)\Big) \;= \\ &=\; \exp\Big(2\vartheta\,t\,\big(\cosh(\dot{\imath}\,s)-1\big)\Big) \;=\; e^{-2\vartheta t}\exp\Big(2\vartheta\,t\big(\cosh(\dot{\imath}\,s)-1\big)\Big).\end{aligned} \tag{3.106}$$

Comparison of the coefficients for individual powers of s through insertion of

$$\cosh(\dot{\imath}s)-1 \;=\; \frac{(\dot{\imath}s)^2}{2!}+\frac{(\dot{\imath}s)^4}{4!}+\frac{(\dot{\imath}s)^6}{6!}+\ldots \;=\; -\frac{s^2}{2!}+\frac{s^4}{4!}-\frac{s^6}{6!}+\ldots$$

yields the individual probabilities:

$$P_n(t) \;=\; I_n(2\vartheta t)\,e^{-2\vartheta t}\,,\; n\in\mathbb{Z}\,. \tag{3.107}$$

where the pre-exponential term is written in terms of modified Bessel functions $I_k(\theta)$ with $\theta = 2\vartheta t$ (for details see [14, p. 208 ff.]), which are defined by

$$\begin{aligned}I_k(\theta) \;&=\; \sum_{j=0}^{\infty}\frac{(\theta/2)^{2j+k}}{j!(j+k)!} \;=\; \sum_{j=0}^{\infty}\frac{(\theta/2)^{2j+k}}{j!\,\Gamma(j+k+1)} \;= \\ &=\; \sum_{j=0}^{\infty}\frac{(\vartheta t)^{2j+k}}{j!(j+k)!} \;=\; \sum_{j=0}^{\infty}\frac{(\vartheta t)^{2j+k}}{j!\,\Gamma(j+k+1)}\,.\end{aligned} \tag{3.108}$$

The probability that the walker is found at his initial location, $n_0\,l$, for example, is given by

$$P_0(t) \;=\; I_0(2\vartheta\,t)\,e^{-2\vartheta\,t} \;=\; \left(1+(\vartheta\,t)^2+\frac{(\vartheta\,t)^4}{4}+\frac{(\vartheta\,t)^6}{36}+\ldots\right)e^{-2\vartheta\,t}$$

Illustrative numerical examples are shown in figure 3.14. It is straightforward to calculate first and second moments from the characteristic function $\phi(s,t)$ by means of equation (2.31) and the result is:

$$\mathrm{E}\big(\mathcal{N}(t)\big) \;=\; n_0 \quad\text{and}\quad \mathrm{var}\big(\mathcal{N}(t)\big) \;=\; 2\vartheta\,(t-t_0)\,. \tag{3.109}$$

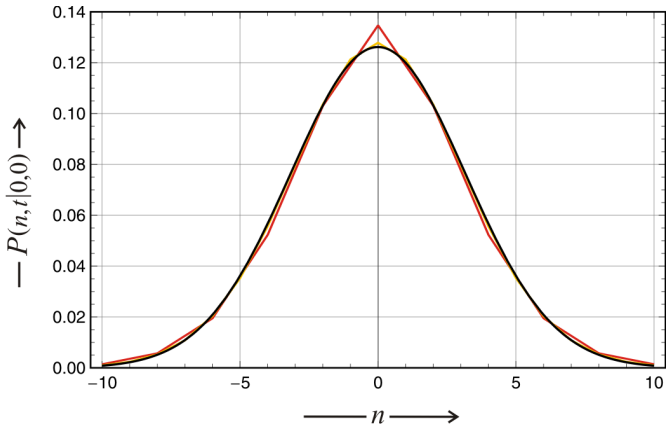


Fig. 3.15 Transition from random walk to diffusion. The figure presents the conditional probabilities $P(n,t|0,0)$ during convergence from a discrete space random walk to diffusion. The black curve is the normal distribution (2.38) resulting from the solution of the stochastic diffusion equation (3.55') with $D = 2\lim_{l\to 0}(l^2\vartheta) = 2$. The yellow curve is the random walk approximation with $l = 1$ and $\vartheta = 1$, the red curve was calculated with $l = 2$ and $\vartheta = 0.25$. Smaller step width of the random walk, $l \leq 0.5$, led to curves that are indistinguishable from the normal distribution. In order to obtain comparable curves, the probability distributions were scaled by a factor $\sigma = l^{-1}$. Choice of other parameters: $t = 5$.

The expectation value is constant and coincides with the starting point of the random walk and the variance increases linearly with time.

The density function $P_n(t)$ allows for straightforward calculation of practically all interesting quantities. For example, we might like to know the probability that the walker reaches a given point at distance $n \cdot l$ from the origin within a predefined time span, which is simply obtained by $P_n(t)$ with $P_n(t_0) = \delta_{n,0}$ (figure 3.14). The probability distribution is symmetric because of the symmetric initial condition $P_n(t_0) = \delta_{n,0}$ and hence $P_n(t) = P_{-n}(t)$. For long times the probability density $P(n,t)$ becomes flatter and flatter and eventually converges to the uniform distribution over the spatial domain. In case $n \in \mathbb{Z}$ all probabilities vanish: $\lim_{t\to\infty} P_n(t) = 0$ for all n.

From random walks to diffusion. In order to derive the stochastic diffusion equation (3.55) we start from a discrete time random walk of a single particle on an infinite one-dimensional lattice where the lattice sites are denoted by $i \in \mathbb{Z}$. Because of its general importance we present two derivations, (i) from the discrete time and space random walk model presented and solved in section 3.1.3.2, and (ii) from the continuous time discrete space random walk (CTRW) discussed in the previous paragraph.

The particle is assumed to be at position i at time t and within a discrete time interval Δt it is obliged to jump to one of the neighboring sites, $i+1$ or $i-1$. This time elapsed between two jumps is called the *waiting time*. Spatial isotropy demands that the probabilities to jump to the right or to the left are the same and equal to one half. The probability to be at site 'i' at time $t+\Delta t$ is therefore given by[37]

$$P_i(t+\Delta t) \;=\; \frac{1}{2}\,P_{i-1}(t) \;+\; \frac{1}{2}\,P_{i+1}(t)\;. \tag{3.9'}$$

Next we make a Taylor expression in time and truncate after the linear term in Δt assuming t is a continuous variable:

$$P_i(t+\Delta t) \;=\; P_i(t)\,\Delta t\,\frac{\mathrm{d}P_i(t)}{\mathrm{dt}} \;+\; \mathcal{O}\big((\Delta t)^2\big)\;.$$

Now we convert the site number into a continuous spatial variable, $i \Rightarrow x$ and $P_i(t) \Rightarrow p\,(x,t)$ and find

$$P_{i\pm1} \;=\; p\,(x,t) \pm \Delta x\,\frac{\partial p\,(x,t)}{\partial x} \;+\; \frac{(\Delta x)^2}{2}\,\frac{\partial^2 p\,(x,t)}{\partial x^2} \;+\; \mathcal{O}\big((\Delta x)^3\big)\;.$$

Here we truncate after the quadratic term in Δx because the terms with the first derivatives cancels, and obtain by insertion into equation (3.9') and omitting residuals

$$\Delta t\,\frac{\partial p\,(x,t)}{\partial t} \;=\; \frac{(\Delta x)^2}{2}\,\frac{\partial^2 p\,(x,t)}{\partial x^2}\;.$$

The next and final task is carrying out the limits to infinitesimal differences in time and space:

$$\lim_{\Delta t\to 0,\Delta x\to 0}\;\frac{(\Delta x)^2}{2\,\Delta t} \;=\; D\;, \tag{3.110}$$

where D is called the diffusion coefficient. According to (3.110) the dimension of D is $[\mathrm{length}^2/\,\mathrm{time} = cm^2 \times sec^{-1}]$. Eventually we obtain the stochastic version of the diffusion equation

$$\frac{\partial p\,(x,t)}{\partial t} \;=\; D\,\frac{\partial^2 p\,(x,t)}{\partial x^2}\;, \tag{3.55'}$$

which is fundamental in physics and chemistry for the description of passive transport by thermal motion (see also equation (3.56) in section 3.2.2.2).

It is also straightforward to consider the continuous time random walk in the limit of continuous space. This is achieved by setting the distance traveled to $x = n \cdot l$ and performing the limit $l \to 0$. For that purpose we start from

[37] It is worth to point out a subtle difference between equations (3.104) and (3.9): The term containing $-2P_i(t)$ is missing in the latter, because moving is obligatory in the discrete time model. The walker is not allowed to take a rest.

the characteristic function of the distribution in x,

$$\phi(s,t) \;=\; \mathrm{E}\Big(e^{\mathrm{i}sx(t)}\Big) \;=\; \Phi(ls,t) \;=\; \exp\Big(2\vartheta\, t\,\big(\cosh(\mathrm{i}\, ls)-1\big)\Big)\;,$$

where ϑ is again the transition probability to neighboring positions per unit time, and make use of the series expansion of the function cosh,

$$\cosh y \;=\; \sum\nolimits_{k=0}^{\infty} \frac{y^{2k}}{(2k)!} \;=\; 1 \,+\, \frac{y^2}{2!} \,+\, \frac{y^4}{4!} \,+\, \frac{y^6}{6!} \,+\, \ldots\;,$$

and take the limit of infinitesimally small steps, $\lim\, l \to 0$,

$$\begin{aligned}\lim\nolimits_{l\to 0}\, \exp\Big(2\vartheta\, t\,\big(\cosh(\mathrm{i}\, ls)-1\big)\, t\Big) \;&=\; \lim\nolimits_{l\to 0}\, \exp\Big(\vartheta\, t\,(-l^2 s^2 \,+\, \ldots)\Big) \;=\\ =\; \lim\nolimits_{l\to 0}\, \exp\left(-s^2\, l^2\, \vartheta\, t\right) \;&=\; \exp\left(-s^2\, D\, t\right)\;,\end{aligned}$$

where we used the definition $D = \lim_{l\to 0}(l^2\vartheta)$ for the diffusion coefficient D (figure 3.15).[38] Since this is the characteristic function of the normal distribution, we obtain for the probability density precisely equation (2.38):

$$p\,(x,t) \;=\; \frac{1}{\sqrt{4\pi D t}}\, \exp\big(-x^2/(4Dt)\big) \tag{2.38}$$

for the sharp initial condition $\lim_{t\to 0} p\,(x,t) = p\,(x,0) = \delta(x)$. We could also have proceeded directly from equation (3.104) and expanded the right-hand side as a function of x up to second order in l, which yields again the stochastic diffusion equation

$$\frac{\partial p\,(x,t)}{\partial t} \;=\; D\, \frac{\partial^2 p\,(x,t)}{\partial x^2}\;, \tag{3.56}$$

where D stands for $\lim_{l\to 0}(l^2\vartheta)$ as before.

Random walks with variable increments. In order to prepare for a discussion of anomalous diffusion that will be presented in the next section 3.2.5 we generalize the 1D continuous time random walk (CTRW) and analyze it from a different perspective [49, 328]. The random variable $\mathcal{X}(t)$ is defined as the sum of previous step increments ξ_k,

$$\mathcal{X}_n(t) \;=\; \sum_{k=1}^{n} \xi_k \quad \text{with} \quad t_n \;=\; \sum_{k=1}^{n} \tau_k\;,$$

and the time t_n is the sum of all earlier waiting times τ_k. This discrete random walk differs from the the previously analyzed case (section 3.1.3.2) by the assumption that both the *jump increments* or jump lengths, $\xi_K \in \mathbb{R}$, and the

[38] The most straightforward way to perform the limit is to introduce a *scaling assumption* using a variable σ such that $l = l_0\sigma$ and $\vartheta = \vartheta_0/\sigma^2$. Then we have $l^2\vartheta = l_0^2\vartheta_0 = D$ and taking the limit $\sigma \to 0$ is trivial.

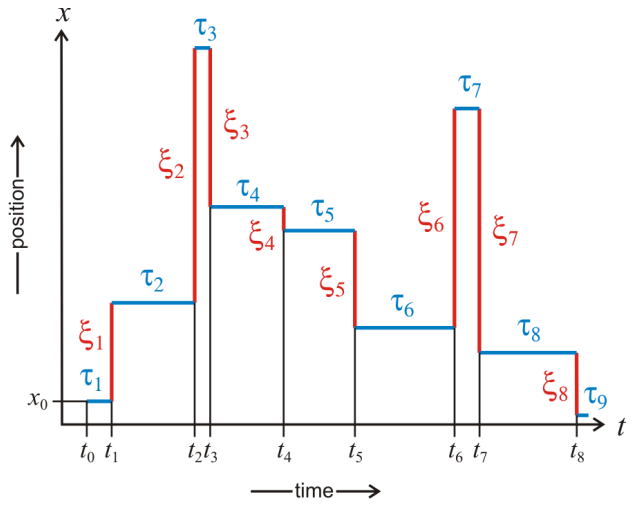


Fig. 3.16 A random walk with variable step sizes. Both, the jump lengths, ξ_k, and the waiting times, τ_k, are assumed to be variable. The jumps occur at times t_1, t_2, ..., and both jump length and waiting times are drawn form the distributions $f(\xi)$ and $w(\tau)$, respectively.

time intervals between two jumps denoted as *waiting times*, $\tau_k \in \mathbb{R}_{\geq 0}$, are variable (figure 3.16). Since jump lengths and waiting times are real quantities the random variable is a continuous as well: $\mathcal{X}(t) \in \mathbb{R}$. The probability at time t_k that the next jump occurs at time $t_k + \Delta t = t_k + \tau_{k+1}$ and that this jump will be of length $\Delta x = \xi_{k+1}$ is given by the joint density function

$$\begin{aligned} &P(\Delta x = \xi_{k+1} \wedge \Delta t = \tau_{k+1} \,|\, \mathcal{X}(t_k) = x_k) \;=\; \varphi(\xi, \tau) \quad \text{with} \\ &\psi(\tau) \;=\; \int_{-\infty}^{+\infty} \mathrm{d}\xi\, \varphi(\xi, t) \quad \text{and} \quad f(\xi) \;=\; \int_0^{\infty} \mathrm{d}\tau\, \varphi(\xi, \tau) \end{aligned} \tag{3.111}$$

being the two marginal distributions. Since $\varphi(\xi, \tau)$ does not depend on time t the process is homogeneous. If we assume that waiting times and jump lengths are independent random variables, the joint density can be factorized:

$$\varphi(\xi, \tau) \;=\; f(\xi) \cdot \psi(\tau)\ . \tag{3.112}$$

In case they were coupled we would have to deal with $\varphi(\xi, \tau) = \varphi(\xi|\tau)\psi(\tau) = \varphi(\tau|\xi)f(\xi)$. Coupling between space and time could arise, for example, from the fact that it is impossible to jump a certain distance within a time span shorter than some minimum time required, but here we assume independence.

In case of Brownian motion or normal diffusion the marginal densities in space and time are Gaussian and exponential distributions modeling normal distributed jump lengths and Poissonian waiting times:

$$f(\xi) \;=\; \frac{1}{\sqrt{4\pi\sigma^2}}\;\exp\!\left(-\frac{\xi^2}{4\sigma^2}\right) \quad\text{and}\quad \psi(\tau) \;=\; \frac{1}{\tau_w}\;\exp\!\left(-\frac{\tau}{\tau_w}\right)\;. \tag{3.113}$$

It is worth recalling that equation (3.113) is sufficient to predict the nature of the probability distributions of $\mathcal{X}_n$ and t_n: Since the spatial increments are iid Gaussian random variables the sum is normal distributed by the central limit theorem (CLT), and since the temporal increments follow an exponential distribution, the probability distribution of the sum is Poissonian. The task is now to express the probability that the random walk is in position x at time t, $p\,(x,t) = P(\mathcal{X}(t) = x|\mathcal{X}(0) = x_0)$, by means of the functions $f(\xi)$ and $\psi(\tau)$. For this goal we calculate first the probability of the walk to arrive at position x at time t under the condition of having been at position z at time ϑ:

$$\eta(x,t) \;=\; p\,(x,t|z,\vartheta) \;=\; \int_{-\infty}^{x}\mathrm{d}z\int_{0}^{\infty}\mathrm{d}\vartheta\;f(x-z)\,\psi(t-\vartheta)\,\eta(z,\vartheta)+\delta(x)\delta(t)\,,$$

with $\psi(t) = 0\,\forall\, t \le \vartheta$. The last term takes into account that the random walk started at the origin, $x = 0$, at time $t = 0$, $p\,(x,0) = \delta(x)$, and defines the initial condition $\eta(0,0) = 1$.

Next we consider the condition that the step $(z,\vartheta) \to (x,t)$ was the last step in the walk until t, and introduce the probability that no step occurred in the time interval $[0,t]$:

$$\Theta(t) \;=\; 1 \;-\; \int_{0}^{t}\mathrm{d}\vartheta\,\psi(\vartheta)\;.$$

Now we can write down the probability density we are searching for

$$p\,(x,t) \;=\; \int_{0}^{t}\mathrm{d}\vartheta\;\Theta(t-\vartheta)\;\eta(x,\vartheta)\;.$$

It is important to realize now that the expression for $\eta(x,t)$ is a convolution of $f(x)$ and $\psi(t)$ and $\eta(x,t)$ with respect to space and time, x and t, respectively, and $p\,(x,t)$ eventually is a convolution of Θ and η with respect to t only.

Making use of the convolution theorem (3.28) that turns convolutions into products we can readily write down the expressions for the probability distributions in Fourier-Laplace space,

$$\begin{aligned}
\hat{\tilde{p}}\,(x,t) &= \hat{\Theta}(u)\cdot\hat{\tilde{\eta}}(k,u) \quad \text{with} \\
\hat{\tilde{\eta}}(k,u) &= \hat{\psi}(u)\tilde{f}(k)+1 \implies \hat{\tilde{\eta}}(k,u) = \frac{1}{1-\tilde{f}(k)\,\hat{\psi}(u)} \quad \text{and} \\
\mathcal{L}\left(\frac{\mathrm{d}\Theta(t)}{\mathrm{dt}}\right) &= \mathcal{L}\left(\delta(t)-\psi(t)\right) \implies u\,\hat{\Theta}(u) = 1-\tilde{f}(k)\;, \\
\hat{\Theta}(u) &= \frac{1-\tilde{f}(k)}{u}\;,
\end{aligned}$$

and obtain the well-known Montroll-Weiss equation [337] named after the American mathematicians Elliot Montroll and George Weiss

$$\hat{\tilde{p}}\,(k,u) = \frac{1-\hat{\psi}(u)}{u}\,\frac{1}{1-\tilde{f}(k)\,\hat{\psi}(u)}\;, \tag{3.114}$$

which provides the desired relation between the increment densities and the probability distribution of the position of the walk as a function of time. What remains to do is to calculate the Fourier (2.30) and Laplace (2.28) transformed increment functions, which are expanded for small values of k and u. The Laplace transform of $\psi(\tau)$ and the Fourier transform of $f(\xi)$ have the asymptotic form

$$\hat{\psi}(u) = \int_0^\infty \mathrm{d}\tau\,\psi(\tau)\,e^{-u\tau} = \frac{1}{1+\tau_w\,u} = 1-\tau_w\,u+\mathcal{O}(u^2) \quad \text{and} \tag{3.115a}$$

$$\sqrt{2\pi}\,\tilde{f}(k) = \int_{-\infty}^{+\infty}\mathrm{d}\xi\,f(\xi)\,e^{i\,k\xi} = e^{-k^2\,\lambda} = 1-\lambda\,k^2+\mathcal{O}(k^4)\;, \tag{3.115b}$$

with $\lambda=\sigma^2/2$ and the diffusion coefficient $D=\lambda/\tau_w=\sigma^2/2\tau_w$. The exponents of the leading terms in the expansions (3.115b), $\alpha=2$, and (3.115a), $\theta=1$ determine the nature of the diffusion process and are called *universality exponents*. Insertion into the Montroll-Weiss equation yields

$$\hat{\tilde{p}}\,(k,u) = \frac{1}{\sqrt{2\pi}}\,\frac{\tau_w}{\tau_w\,u+\lambda\,k^2} = \frac{1}{\sqrt{2\pi}}\,\frac{1}{u+D\,k^2} \tag{3.115c}$$

As expected consecutive inverse transformations yield the density distribution $p\,(x,t)$ of the Wiener process (3.61):

$$\begin{aligned}
\mathcal{F}^{-1}\left(\frac{1}{\sqrt{2\pi}}\,\frac{1}{u+D\,k^2}\right) &= \frac{1}{\sqrt{4Du}}\,e^{-\sqrt{\frac{u}{D}}\,|x|} \quad \text{and} \\
\mathcal{L}^{-1}\left(\frac{1}{\sqrt{4Du}}\,e^{-\sqrt{\frac{u}{D}}\,|x|}\right) &= \frac{1}{\sqrt{4\pi Dt}}\,e^{-\frac{x^2}{4Dt}}\;,
\end{aligned}$$

with D being the diffusion coefficient and $x_0 = 0$, $t_0 = 0$ as initial conditions. It is a good exercise to show that inverting the order of the transformations yields the same result:

$$\mathcal{L}^{-1}\left(\frac{1}{\sqrt{2\pi}}\,\frac{1}{u\,+\,D\,k^2}\right) \;=\; \frac{1}{\sqrt{2\pi}}\,e^{-D\,k^2 t} \quad \text{and}$$
$$\mathcal{F}^{-1}\left(\frac{1}{\sqrt{2\pi}}\,e^{-D\,k^2 t}\right) \;=\; \frac{1}{\sqrt{4\pi D t}}\,e^{-\frac{x^2}{4Dt}}\;.$$

If one were only interested in the solution for normal distribution, the derivation of the solution presented here would be a true case of *overkill*. We shall, however, extend the analysis to anomalous diffusion with generalized exponents $0 < \alpha \leq 2$ and $0 < \theta \leq 1$, which are non-integer quantities and thus lead into the realm of fractals (section 3.2.5).

Before we give an interpretation of the two expansions we visualize the meaning of the two transformed variables u and k: The exponents, $u \cdot \tau$ and $\xi \cdot k$ are dimensionless quantities and hence the dimensions of u and k are reciprocal time, $[t^{-1}]$, and reciprocal length, $[l^{-1}]$ respectively. The values $u = 0$ and $k = 0$ of the transformed variables refer to infinite time and infinite space and accordingly, expansions around these points are valid for long times and large distances. Commonly, the problem specific properties are dominant at short times and small distances and universal behaviour is expected to be found on the opposite ends of the time scale and space as expressed by vanishing u and k. Both transformed probability distributions in equations (3.115a) and (3.115b) are given in expressions that allow for direct readout of the so-called *universality exponents*, which are $\alpha = 2$ for the spatial density $\tilde{f}(k)$ and $\theta = 1$ for the temporal density $\hat{\psi}(u)$. Random walks can be classified by the variance of jump lengths, $\mathrm{var}(\xi)$, and by the expectation value of waiting times, $\mathrm{E}(\tau)$:

(i) the *variance of the jump length* $\left\langle \xi^2 \right\rangle = 2\sigma^2 = \int_{-\infty}^{+\infty} \mathrm{d}\xi\, \xi^2\, f(\xi)$.[39] and

(ii) the *characteristic* or *mean waiting time* $\langle \tau \rangle = \tau_w = \int_0^{\infty} \mathrm{d}\tau\, \tau\, w(\tau)$,

which are both finite quantities that do not diverge in the integration limits $\xi \to \infty$ and $\tau \to \infty$ in contrast to Lévy processes with $0 < \alpha < 2$ and $0 < \theta < 1$, which will be discussed in section 3.2.5. As a matter of fact any pair of probability density functions with finite τ_w and σ^2 leads to the same asymptotic result and this is a beautiful manifestation of the central limit theorem (section 2.4.2): In the inner part of the transformed densities all representatives of the universality class of CTRWs with finite mean waiting times and positional variances fulfil equations (3.115a) and (3.115b) and the individuality of the densities comes into play only within the higher order terms $\mathcal{O}(\tau^2)$ and $\mathcal{O}(k^4)$.

[39] As in the previous examples we assume that the random walk is symmetric and started at the origin. Then the expectation value of the location of the particle stays at the origin and we have $\langle \xi \rangle = 0$, $\langle \xi \rangle^2 = 0$, and hence $\mathrm{var}(\xi) = \langle \xi^2 \rangle$.

Finally, we mention a feature that will be brought up again and generalized in the next section 3.2.5: the Wiener process is self-similar. A stochastic process is self-similar with Hurst index H, named after the British hydrologist Harold Edwin Hurst, if the two processes

$$(\mathcal{Y}(a\,t), t \geq 0) \quad \text{and} \quad (a^H \mathcal{Y}(t), t \geq 0)$$

with the same initial condition $\mathcal{Y}(0) = 0$ have the same finite-dimensional distribution for all $a \geq 0$. Expressed in popular language if you look on a self-similar process with a magnifying glass it looks the same as without the magnifier no matter how large the magnification factor is. The expectation value of a generalized Brownian processes $\mathcal{B}(t)$ at two different times, t_1 and t_2, and with Hurst index $0 \leq H \leq 1$ is

$$\mathrm{E}\big(\mathcal{B}_H(t_1)\mathcal{B}_H(t_2)\big) \;=\; \frac{1}{2}\big(|t_2|^{2H} + |t_2|^{2H} - |t_2 - t_1|^{2H}\big)\ .$$

For $H = 1/2$ we are dealing with conventional Brownian motion or Wiener processes where the expectation value $\mathrm{E}\big(\mathcal{B}_H(t_1)\mathcal{B}_H(t_2)\big)$ is zero and increments at different times are uncorrelated as we showed in section 3.2.2.2.

3.2.5 Lévy processes

Lévy processes can be understood as comprehensive generalizations of random walks to continuous time and in this sense they represent the simplest class of stochastic processes whose trajectories consist of continuous motion interrupted by discontinuous jumps of random size occurring at random times. Lévy processes were defined in precise mathematical terms and analyzed in detail by the famous French mathematician Paul Lévy. They constitute a core theme of financial mathematics [11, 111, 392] and they are indispensable constituents of every course in theoretical economy. Many stochastic processes from other fields and also from science fall into this class: From the examples of stochastic processes we have already encountered here, Brownian motion (section 3.2.2.2), the Poisson process (section 3.2.2.4), the random walk (section 3.2.4), and the Cauchy process (section 3.1.3.5) are Lévy processes. Among other applications Lévy processes are used in the mathematical theory of anomalous diffusion [49, 328] and other forms of fractional kinetics, and in Lévy flights based on probability densities with heavy tails, which were applied, for example, in behavioral biology to modeling foraging strategies of animals.

We are also interested here in Lévy processes, because they allow for a general analytic treatment combining all three classes of processes appearing in the differential Chapman-Kolmogorov equation (dCKE): drift, diffusion and jump. This is possible because of the simplifying assumption that all

random variables $\mathcal{X}(t)$ are independent and identically distributed (iid) and all increments $\mathcal{Z} = \Delta\mathcal{X}(t)$ depend only on Δt and not on t explicitly. The time dependence is then restricted to the probability densities $p\,(x,t)$, and the functions $A(x,t)$ and $B(x,t)$ as well as the transition probabilities $W(x|z,t)$ are strictly time independent.

A Lévy process $\mathcal{X} = (\mathcal{X}(t), t \geq 0)$ is a stochastic process that satisfies the following four properties:

(i) the random variable $\mathcal{X}(t)$ has independent increments as expressed by the property that the variables $\mathcal{Z}_k = \mathcal{X}(t_k) - \mathcal{X}(t_{k-1})$ with $k = 1, 2, \ldots$ are statistically independent,

(ii) the increments $\mathcal{Z}_k$ of the random variable $\mathcal{X}(t)$ are stationary in the sense that the probability distributions of the increments $\mathcal{Z}_k$ depend only on the length of the time interval $\Delta t = t_k - t_{k-1}$ but do not depend explicitly on time t, and increments on equal time intervals are identically distributed,

(iii) the process starts at the origin, $\mathcal{X}_0 = 0$, with probability one, and

(iv) the trajectory of the random variable $\mathcal{X}(t)$ is at least piecewise *stochastically continuous* in the sense that it fulfils the relation

$$\lim_{t_1 \to t_2} P(|\mathcal{X}(t_2) - \mathcal{X}(t_1)| > a) = 0$$

for all $a > 0$ and for all $t_2 - t_1 \geq 0$.

The conditions (i), (ii), and (iii) simplify the general dCKE substantially. Conditions (ii) in particular allows the substitution of functions by parameters:

$$\begin{aligned} &A(x,t) \;\text{ is replaced by }\; a\,, \;\; B(x,t) \;\text{ is replaced by }\; \frac{1}{2}\,\sigma^2\,, \\ &\text{and }\; W(x|z,t) \;\text{ is replaced by }\; w(x-z)\,, \end{aligned} \tag{3.116}$$

where $w(x-z)$ is a *transition function* replacing the transition matrix. For the initial condition $p\,(x,t_0) = \delta(x_0)$ the dCKE has the form[40]

$$\begin{aligned} \frac{\partial p\,(x,t)}{\partial t} \;=\; &-\, a\,\frac{\partial p\,(x,t)}{\partial x} \,+\, \frac{1}{2}\,\sigma^2\,\frac{\partial^2 p\,(x,t)}{\partial x^2} \,+ \\ &+ \int \mathrm{d}z\,\big(w(x-z)\,p\,(z,t) - w(z-x)\,p\,(x,t)\big)\;. \end{aligned} \tag{3.117}$$

Lévy processes are thus fully characterized by the Lévy-Khinchin triplet (a, σ^2, w) which is named after Paul Lévy and the Russian mathematician

[40] For Lévy processes in general it will be necessary to replace the integral by a principal value integral because they may lead to a singularity at the origin, $\lim_{z \to x} w(x-z) = \infty$, which is prohibitive for conventional integration.

Aleksandr Khinchin. As follows from condition (ii) a Lévy process is a homogeneous Markov process.

The replacement of the functions $A(x,t)$ and $B(x,t)$ by the constants a and $\frac{1}{2}\sigma^2$, and the elimination of time from the jump probability $W(z|x,t)$ has remarkable analogy to linearity in deterministic dynamical systems. Indeed, the Liouville equation corresponding to the dCKE (3.117) gives rise to a linear time dependence

$$\frac{\partial p\,(x,t)}{\partial t} \;=\; -a\,\frac{\partial p\,(x,t)}{\partial x} \quad\Longrightarrow\quad \frac{\mathrm{d}x}{\mathrm{dt}} \;=\; a \;\;\text{and}\;\; x(t) = x(0) + at = at\;,$$

the diffusion part is a Wiener process with a linearly growing variance

$$\frac{\partial p\,(x,t)}{\partial t} \;=\; \frac{1}{2}\sigma^2\,\frac{\partial^2 p\,(x,t)}{\partial x^2} \quad\Longrightarrow\quad \mathrm{var}\big(\mathcal{X}(t)\big) \;=\; \sigma^2\,\mathrm{var}\big(\mathcal{W}(t)\big) \;=\; \sigma^2\,t\;,$$

and the jumps have time independent transition probabilities where the analogy to a linear process is a little bit far-fetched.

Characteristic functions of Lévy processes. Equation (3.117) for a the Lévy process starting at $t_0 = 0$ from $p\,(x,0) = \delta(x_0)$ is solved now by means of the characteristic function as is defined in section 2.2.3

$$\frac{\partial}{\partial t}\phi(s,t) \;=\; \frac{\partial}{\partial t}\int_{-\infty}^{+\infty}\mathrm{d}x\,e^{\imath\!\imath\,sx}\,p\,(x,t) \;=\; \int_{-\infty}^{+\infty}\mathrm{d}x\,e^{\imath\!\imath\,sx}\,\frac{\partial p\,(x,t)}{\partial t}\;.$$

Insertion into equation (3.117) and application of integration by parts to the first two terms (see section 3.2.2.2) yields the differential equation

$$\frac{\partial\phi(s,t)}{\partial t} \;=\; \Big(\imath\!\imath\,a\,s \;-\; \frac{1}{2}\,\sigma^2 s^2 \;+\; \mathrm{III}\Big)\,\phi(s,t)\;.$$

The third term in the parentheses, III, is calculated with a little trick: We substitute $z - x \;\Rightarrow\; u$, apply $\mathrm{d}z = \mathrm{d}u$ and find for the second summand

$$\int\!\mathrm{d}z\,w(z-x)\int\!\mathrm{d}x\,e^{\imath\!\imath sx}p\,(x,t) = \int\!\mathrm{d}u\,w(u)\int\!\mathrm{d}x\,e^{\imath\!\imath sx}p\,(x,t) = \int\!\mathrm{d}u\,w(u)\,\phi(s,t)\,,$$

whereas the first summand is calculated by means of a shift in the variable:

$$\int\!\mathrm{d}x\int\!\mathrm{d}z\,w(x-z)\,e^{\imath\!\imath sx}p\,(z,t) = \int\!\mathrm{d}(-u)\,w(-u)\int\!\mathrm{d}z\,e^{\imath\!\imath s(z-u)}p\,(z,t) =$$

$$= \int\!\mathrm{d}(-u)\,w(-u)\,e^{\imath\!\imath s(-u)}\int\!\mathrm{d}z\,e^{\imath\!\imath sz}\,(z,t) = \int\!\mathrm{d}u\,w(u)\,e^{isu}\;\phi(s,t)\;.$$

Collecting all terms yields the differential equation for the characteristic function:

$$\frac{\partial\phi(s,t)}{\partial t} \;=\; \Big(\dot{\imath}as \;-\; \frac{1}{2}\sigma^2 s^2 \;+\; \int \mathrm{d}u\, w(u)(e^{\dot{\imath}su} - 1)\Big)\,\phi(s,t)\;, \qquad (3.118)$$

which can be readily solved. A principal value integral is introduced in order to make sure that the solution is not jeopardized by a possible singularity of $w(u)$ [157, pp. 248-252] leads to

$$\begin{aligned}\phi(s,t) \;&=\; \int_{-\infty}^{+\infty} \mathrm{d}x\, e^{\dot{\imath}sx}\, p\,(x,t) \;= \\ &=\; \exp\!\bigg(\Big(\,\dot{\imath}\,a\,s \;-\; \frac{1}{2}\,\sigma^2 s^2 \;+\; \fint_{-\infty}^{+\infty} \mathrm{d}u\,(e^{\dot{\imath}su} - 1)\,w(u)\Big)\,t\bigg)\;.\end{aligned} \qquad (3.119)$$

The density of the Lévy process can be obtained, for example, by inverse Fourier transform. Often the factor in the exponent of the exponential function is denoted as the characteristic exponent

$$\psi(s) \;=\; \frac{\ln\phi(s,t)}{t} \;=\; \dot{\imath}\,a\,s \;-\; \frac{1}{2}\,\sigma^2 s^2 \;+\; \fint_{-\infty}^{+\infty} \mathrm{d}u\,(e^{\dot{\imath}su} - 1)\,w(u)\,. \qquad (3.119')$$

In practice it is often appropriate to circumvent the sophistication caused by a possible singularity in the integral of equation (3.119) at $u = 0$ as it is done, for example, in the Lévy-Khinchin formula

$$\begin{aligned}&\fint_{-\infty}^{+\infty} \mathrm{d}u\, w(u)\,\big(e^{\dot{\imath}su} - 1\big) \;\equiv \\ &\equiv\; \lim_{\epsilon\to 0}\left(\int_{-\infty}^{-\delta(\epsilon)} \mathrm{d}u\, w(u)\big(e^{\dot{\imath}su} - 1\big) \;+\; \int_{\epsilon}^{+\infty} \mathrm{d}u\, w(u)\big(e^{\dot{\imath}su} - 1\big)\right) \;= \\ &=\; \dot{\imath}s\; a_{\mathcal{L}} \;+\; \int_{-\infty}^{+\infty} \mathrm{d}u\,(e^{\dot{\imath}su} - 1 - \dot{\imath}su\;\mathbf{1}_{|u|<1})\,w(u) \quad \text{with} \\ &\qquad \dot{\imath}s\; a_{\mathcal{L}} \;=\; \lim_{\epsilon\to 0}\left(\int_{-1}^{+\delta(\epsilon)} \mathrm{d}u\,\dot{\imath}su\; w(u) \;+\; \int_{\epsilon}^{+1} \mathrm{d}u\,\dot{\imath}su\; w(u)\right) \;+\; \kappa\;.\end{aligned} \qquad (3.120)$$

Herein $\delta(\epsilon)$ is a function controlling the asymmetric interval used in the calculation of the limit in the evaluation of the principal value integral is of the form $\delta(\epsilon) = C\epsilon + \kappa$ that fulfils $\delta(\epsilon) \to 0$ if $\epsilon \to 0$, where C is a factor to be derived from the transition function $w(u)$ and κ is the integration constant.

With this modifications the characteristic exponent of a general Lévy process becomes

$$\psi(s) \;=\; \dot{\imath}(a + a_{\mathcal{L}})s - \frac{1}{2}\,\sigma^2 s^2 + \fint_{-\infty}^{+\infty} \mathrm{d}u\,(e^{\dot{\imath}su} - 1 - \dot{\imath}su\,\mathbf{1}_{|u|<1})\,w(u)\,, \qquad (3.119'')$$

where the evaluation of the principle value integral by residue calculus is shifted into the calculation of the parameter $a_{\mathcal{L}}$.

In the following paragraphs we present a few examples of Lévy processes.

Poisson processes. The conventional Poisson process and two modifications of it are discussed as examples of simple Lévy processes. As said before the Poisson process is a Lévy process with the parameters $a = \sigma = 0$ and the transition function $w(u) = \gamma\,\delta(u-1)$. The solution is a Poisson distribution with the parameter $\lambda = \gamma t$:

$$P_n(t) \;=\; e^{-\gamma t}\,\frac{(\gamma t)^n}{n!}\;. \tag{3.88}$$

The parameter γ is denoted as the intensity of the process, and it represents the inverse of the mean time between two jumps.

The compensated Poisson process is an other example of a Lévy process. The stochastic growth of the Poisson process is compensated by a linear deterministic term, and the two parameters and the transition function are: $a = -\gamma$, $\sigma = 0$, and $w(u) = \gamma\,\delta(u-1)$. The process is described by the random variable $\mathcal{X}(t) = \mathcal{Z}(t) - \gamma t$ with the expectation value where $\mathrm{E}\big(\mathcal{X}(t)\big) = 0$, where $\mathcal{Z}(t)$ is a Poisson process. Accordingly the compensated Poisson process is a martingale (section 3.1.3.2).

An important generalization of the conventional Poisson process is the *compound Poisson process*, which is a Poisson process with variable step sizes expressed as random variables $\mathcal{X}_k$ drawn from a probability density $w(u)/\gamma$:

$$f(u)\,\mathrm{d}u \;=\; P\big(u < \mathcal{X}_k < u + \mathrm{d}u\big) \;=\; \frac{w(u)}{\gamma}\,\mathrm{d}u\;. \tag{3.121}$$

The transition function is assumed to be normalizable

$$\gamma \;=\; \int_{-\infty}^{\infty} w(u)\,\mathrm{d}u \;<\; \infty\;,$$

where γ is again the intensity of the process. The number of events in a compound Poisson process that happened until time t is described by the random variable $\mathcal{Z}(t) = \sum_{k=1}^{n(t)} \mathcal{X}_k$. The Poisson process and the compounds Poisson process are the one-sided analogues of the constant and variable step size random walks (figure 3.16).

Wiener processs. The Wiener process follows trivially from equation 3.117 by choosing $a = 0$, $\sigma = 1$, and setting zero probability for jumps, which leads to the characteristic function

$$\phi(s,t) \;=\; e^{-\frac{1}{2}\,s^2 t}\;.$$

This is the characteristic function for the Wiener process (3.60) with $w_0 = 0$ and $t_0 = 0$. Interestingly, we shall see that the Wiener process can be

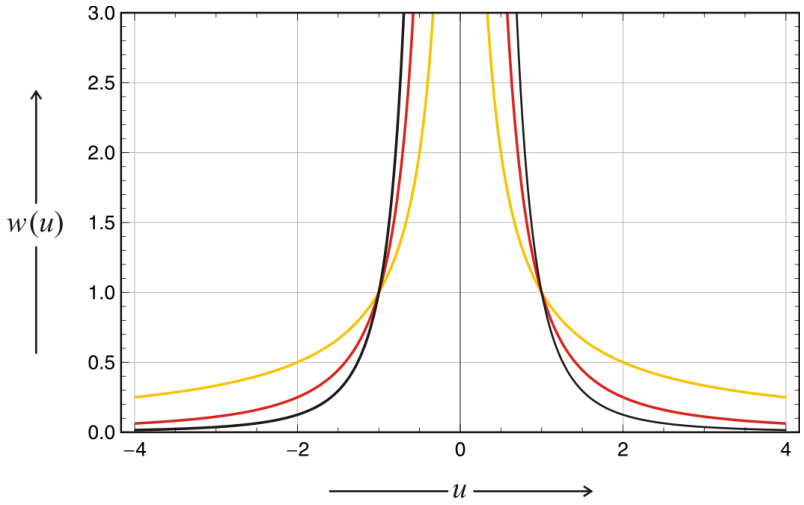


Fig. 3.17 Transition functions of Pareto processes. The transition functions $w(u) = |u|^{-(\alpha+1)}$ of the Pareto processes with $\alpha = 0$ (yellow), $\alpha = 1$ (red) and $\alpha = 2$ (black) are plotted against the variable u. The curve for $\alpha = 2$ is the reference corresponding to normal diffusion. All curves with $2 > \alpha \geq 0$ have heavier tails and this implies a larger probability for longer jumps.

obtained also from (3.117) with $a = \sigma = 0$ and a special transition function $w(u) \propto |u|^{-(\alpha+1)}$ with $\alpha = 2$ (see next paragraph).

Pareto processes. Pareto or Paretian processes are special pure jump Lévy processes with $a = \sigma = 0$ and a transition function of the type

$$w(u) \;=\; \begin{cases} \vartheta_- \cdot |u|^{-(\alpha+1)} & \text{for} \quad -\infty < u < 0 \\ \vartheta_+ \cdot \; u^{-(\alpha+1)} & \text{for} \quad 0 < u < \infty \end{cases} , \tag{3.122}$$

with $0 < \alpha < 2$. In figure 3.17 we show the transition functions for Pareto processes with the singularity at $u = 0$ as in (3.122). We are now in a position to choose an appropriate function for the evaluation of the principal value integral by means of the Lévy-Khinchin formula as expressed in (3.120): $\delta(\epsilon) = \left(\vartheta_+ \epsilon^{1-\alpha} + \kappa)/\vartheta_-\right)^{1/(\alpha-1)}$. Apparently, the case $\alpha = 1$ cannot be handled in this way but there is the possibility of a direct integration without using the Lévy-Khinchin formula.

For Pareto processes the principal value integral can be calculated directly by means of Cauchy's integration through analytic continuation in the complex plane, $z = u + \dot{\imath}v = |z|e^{\dot{\imath}\vartheta}$, and residue calculus [14, chs. 6 and 7]:

$$\oint_\gamma f(z)\mathrm{d}z \;=\; 2\pi\dot{\imath}\,\mathrm{Res}\big(f(z),z_0\big) \;=\; 2\pi\dot{\imath}\;a_{-1}\;,$$

$$\;=\; 2\pi\dot{\imath}\;\lim_{z\to z_0}\big((z-z_0)f(z)\big) \;=\;, \text{ and} \tag{3.123a}$$

$$\;=\; 2\pi\dot{\imath}\,\frac{1}{(m-1)!}\lim_{z\to z_0}\frac{\mathrm{d}^{m-1}}{\mathrm{d}z^{m-1}}\Big((z-z_0)^m f(z)\Big)\;, \tag{3.123b}$$

where γ is a closed contour encircling the (isolated) pole at $z=z_0$ in a region where $f(z)$ is analytical, $\mathrm{Res}\big(f(z),z_0\big)$ is the residue of $f(z)$ at this pole, and a_{-1} is the coefficient of $(z-z_0)^{-1}$ in a Laurent series[41]

$$f(z) \;=\; \sum_{n=-\infty}^{\infty} a_n\,(z-z_0)^n \;\text{ with }\; a_n \;=\; \frac{1}{2\pi\dot{\imath}}\oint_\gamma \frac{f(z)}{(z-z_0)^n}\;. \tag{3.123c}$$

If $f(z)$ has a pole of order m at $z=z_0$ all coefficients a_n for $n<-m<0$ with $n\in\mathbb{Z}$ are zero and $a_{-m}\neq 0$ defines the first non-vanishing term of the series.

According to (3.122) the transition function $w(u)$ has a pole of order $\alpha+1$ at $u=u_0=0$ and since α need not be an integer the analysis of Pareto processes opens the door into the world of fractals. It is worth noticing that the Γ-function as well as the factorials adopt an infinite number of factors for non-integer arguments, $\alpha!=\alpha\cdot(\alpha-1)\cdot(\alpha-2)\cdot(\alpha-3)\ldots$. Evaluation of the characteristic exponent yields for $\kappa=0$:

$$\psi(s) \;=\; |s|^\alpha\,\Gamma(-\alpha)\left((\vartheta_+ + \vartheta_-)\,\cos\frac{\alpha\pi}{2} \;-\; \dot{\imath}\frac{s}{|s|}(\vartheta_+ - \vartheta_-)\,\sin\frac{\alpha\pi}{2}\right) =$$

$$=\; -\,|s|^\alpha\,\chi\,\left(1-\dot{\imath}\beta\,\frac{s}{|s|}\,\omega(s,\alpha)\right) \;\text{ with}$$

$$\omega(s,\alpha) \;=\; \begin{cases}\tan\frac{\alpha\pi}{2}\,, & \text{if } \alpha\neq 1\,, 0<\alpha<2\,,\\ -\frac{2}{\pi}\,\ln|s|\,, & \text{if } \alpha=1\;,\end{cases}$$

$$\beta \;=\; \frac{\vartheta_+ - \vartheta_-}{\vartheta_+ + \vartheta_-} \;\text{ and }\; \chi=\gamma^\alpha=-(\vartheta_+ + \vartheta_-)\,\Gamma(-\alpha)\cos(\alpha\pi/2)\;,$$

and for the characteristic function of Pareto processes we finally obtain:

$$\phi(s,t) \;=\; \exp\Big(-\,|s|^\alpha\,\gamma^\alpha\,\big(1\;-\;\dot{\imath}\beta\,\frac{s}{|s|}\,\omega(s,\alpha)\big)\,t\Big)\;. \tag{3.124}$$

The parameter β determines the symmetry of the transition function and the process: $\beta=0$ implies invariance with respect to inversion at the the origin, $u\to -u$ or $s\to -s$, respectively, and $\beta\neq 0$ expresses the skewness of the characteristic function and the density. Distributions with $\beta=\pm 1$ are called

[41] The Laurent series is an extension of the Taylor series to negative powers of $(z-z_0)$ named in honor of the French mathematician Pierre Alphonse Laurent.

extremal and for $\alpha < 1, \beta \pm 1$ they are one-sided (see for example the Lévy distribution in section 2.5.7 and figure 3.18).

Previously we encountered already the two symmetric processes – $\beta = 0$, $\vartheta_+ = \vartheta_- = \vartheta$, and $\gamma^\alpha = \chi = -2\vartheta\,\Gamma(-\alpha)\cos(\alpha\pi/2)$ – with $\alpha = 1$ and $\alpha = 2$ for which analytical probability densities are available: (i) the Cauchy process ($\alpha = 1$ and $\gamma = \vartheta\pi$)[42]

$$\phi(s,t) \;=\; \exp(-\,\pi\,|s|\,\vartheta t) \;\text{ and }\; p\,(x,t) \;=\; \frac{1}{\pi}\,\frac{\vartheta t}{(\vartheta t)^2 \,+\, x^2}\;,$$

and (ii) the Wiener Process ($\alpha = 2$ and $\gamma = \vartheta$)[43]

$$\phi(s,t) \;=\; \exp(-s^2\vartheta t) \;\text{ and }\; p\,(x,t) \;=\; \frac{1}{\sqrt{4\pi\,\vartheta t}}\,\exp\!\big(x^2/(4\,\vartheta t)\big)\;,$$

where the probability densities were normalized after the Fourier transform. As seen already from equation (3.124) the role of the parameter γ or also θ in the symmetric case is only a scaling of the times axis: $t \to \tau = \theta t$.

Interpretation of sigularities in the transition functions. Lévy processes and Pareto processes in particular are defined also for cases where the transition functions have singularities at the origin $u = u_0 = 0$. In other words, we are considering examples with $\lim_{u\to 0} w(u) = \infty$. How can we visualize such a situation? Apparently the condition modeled by the singularity implies the occurrence of shorter and shorter jumps at higher and higher rates until we are dealing in the limit $u \to 0$ with an infinite number of steps of infinitesimal size. Actually taking this limit is not unfamiliar to us, because in the transition from random walk to diffusion precisely the same problem arose and was solved straightforwardly. We mention again that diffusion appears twice in Lévy processes: (i) in the diffusion term of equation (3.117) and (ii) at the singularity of the transition function $w(u)$. As shown in figure 3.17 a higher order, $m = \alpha+1$, of the pole at $u = 0$ is accompanied by a *broader* singularity and a less heavy tail. The parameter α is confined to the range $0 < \alpha < 2$ and we can expect more diffusion-like behavior the closer α approaches the value $\alpha = 2$, which is the limit of normal diffusion . Smaller values of α result in higher probabilities for longer jumps (see Lévy flights).

Although the characteristic function can be written down for any Lévy process, the probability density need not be expressible in terms of analytic functions. Three examples of Lévy processes where full analytical access is possible are shown in (figure 3.18): (i) the normal distribution (section 2.3.3)

---

[42] Hereby we are using the easy to check relations $\lim_{\alpha\to 1}\Gamma(-\alpha) = \pm\infty$ but $\lim_{\alpha\to 1}\Gamma(-\alpha)\cos(\pi\alpha/2) = -\pi/2$.

[43] Although the value $\alpha = 2$ leads to divergence in the regular derivation, applying $\alpha = 2$, $\beta = 0$, and $\chi = \vartheta = D$ yields the probability density of the normal diffusion process.

with $\alpha = 2$, $\beta = 0$ and $\gamma = 1/2$,[44] the Cauchy distribution (section 2.5.6) with $\alpha = 1$ and $\beta = 0$,[44] and the Lévy distribution (section 2.5.7) with $\alpha = 1/2$ and $\beta = 1$. For work in practice this is hardly a restriction since numerical computation allows for handling all cases and most mathematics packages contain fast routines for Lévy and Pareto distributions.

Infinite divisibility and stability. The property of infinite divisibility is defined for classes of probability densities $p\,(x)$ and requires that a random variable $\mathcal{S}$ with this density, $p\,(x)$, can be partitioned into any arbitrary number n with $n \in \mathbb{N}_{>0}$ of independent and identically distributed (iid) random variables such that all individual variables $\mathcal{X}_k$, the sum $\mathcal{S}_n = \mathcal{X}_1 + \mathcal{X}_2 + \ldots + \mathcal{X}_n$, and all possible partial sums have the same probability density $p\,(x)$. Lévy processes are homogeneous Markov processes and they are infinitely divisible. In general, however, the probability distributions of the individual parts $\mathcal{X}_k$ will be different and different from the density $p\,(x)$.

We define: a random variable, $\mathcal{X}$, has a stable distribution if any linear combination of two independent copies of this variable, $\mathcal{X}_1$ and $\mathcal{X}_2$, fulfils the same distribution up to a shift in location (μ) and a change in the change in the scale parameter being the standard deviation,

$$a\,\mathcal{X}_1 \,+\, b\,\mathcal{X}_2 \;\stackrel{d}{=}\; c\,\mathcal{X} \,+\, d\;, \tag{3.125}$$

wherein a and b are positive constants, c is some positive number dependent on a, b and the summation properties of $\mathcal{X}$, $d \in \mathbb{R}$, and the symbol 'd' above the equals sign means equality in distribution. A Lévy process, $(\mathcal{X}_t, t \geq 0)$ is called stable if every random variable $\mathcal{X}_t$ has a stable distribution [347]. *Stability* or *stability in the broad sense* is to be distinguished from *strict stability* or *stability in the narrow sense* in which case the equality 3.125 holds with $d = 0$ for all choices of a and b. A random variable is called *symmetric stable* if it is stable and symmetrically distributed around zero as expressed by $\beta = 0$ or $\mathcal{X} \stackrel{d}{=} -\mathcal{X}$.

Stability and strict stability of the normal distribution $\mathcal{N}(\mu, \sigma)$ can be easily demonstrated by means of CLT:

$$\begin{aligned} \mathcal{S}_n \;&=\; \sum_{i=1}^{n} \mathcal{X}_i \;\text{ with }\; \mathrm{E}(\mathcal{X}_i) \,=\, \mu\,,\; \mathrm{var}(\mathcal{X}_i) \,=\, \sigma^2 \;\forall\; i = 1, \ldots, n \\ \mathrm{E}(\mathcal{S}_n) \;&=\; n\,\mu \;\text{ and }\; \mathrm{var}(\mathcal{S}_n) \;=\; (n\,\sigma)^2 \;\forall\; i = 1, \ldots, n\;. \end{aligned} \tag{3.126}$$

From the two equations (3.125) and (3.126) follow the conditions for the constants a, b, c, and d:

[44] The requirement for the derivation of the formula of the characteristic function of Pareto processes was $0 < \alpha < 2$, which does not include the case $\alpha = 2$ and indeed the expression for γ diverges for $\alpha \to 2$: $\Gamma(-2)\cos\pi = \pm\infty$. For $\alpha = 1$ we find $\Gamma(-1) = \infty$, $\Gamma(-1)\sin(\pi/2) = \infty$ but $\lim_{\alpha\to 1}\Gamma(-\alpha)\cos(\alpha\pi/2) = -\pi/2$, which allows for the derivation of the characteristic function of the Cauchy process.

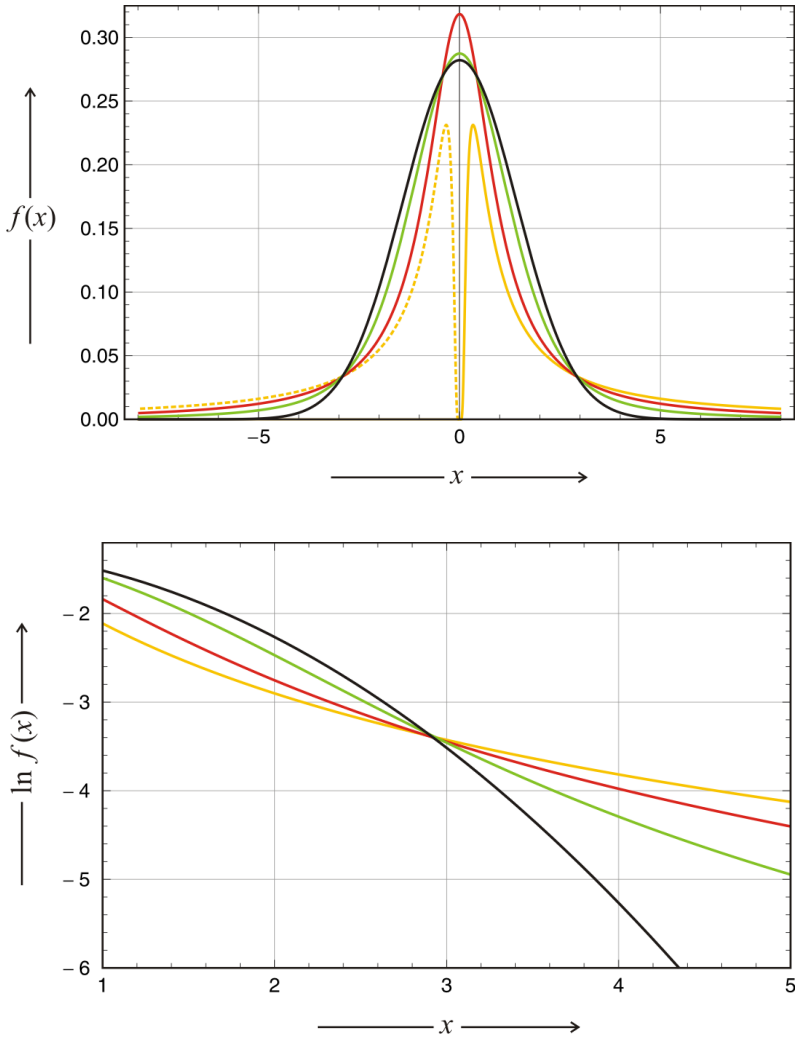


Fig. 3.18 A comparison of stable Pareto probability densities. In the upper part of the figure four different stable distributions with characteristic exponents $\alpha = 1/2$ (yellow), 1 (red), 3/2 (green), and 2 (black) are compared. For $\alpha < 1$ symmetric distributions ($\beta = 0$) are not stable and therefore we show the two extremal distributions with $\beta = \pm 1$ for the Lévy distribution ($\alpha = 1/2$). The lower part presents a log-linear plot of the densities against the position x. Within a small interval around $x = 2.9$ the curves for the individual probability densities cross and illustrate the increase in the probabilities for longer jumps.

$$\mu(a\mathcal{X}) = a\,\mu(\mathcal{X})\,,\ \mu(b\mathcal{X}) = b\,\mu(\mathcal{X})\,,\ \mu(c\mathcal{X}+d) = c\,\mu(\mathcal{X}) + d \ \Rightarrow$$
$$\Rightarrow\ d = (a+b-c)\,\mu$$
$$\mathrm{var}(a\mathcal{X}) = (a\sigma)^2\,,\ \mathrm{var}(b\mathcal{X}) = (b\sigma)^2\,,\ \mathrm{var}(c\mathcal{X}+d) = (c\sigma)^2 \ \Rightarrow$$
$$\Rightarrow\ c^2 = a^2 + b^2\ .$$

The two conditions $d = (a+b-c)\,\mu$ and $c = \sqrt{a^2+b^2}$ with $d \neq 0$ are readily fulfilled for pairs of arbitrary real constants $a, b \in \mathbb{R}$ and accordingly, the normal distribution $\mathcal{N}(\mu,\sigma)$ is stable. Strict stability, on the other hand, requires $d = 0$ and this can be fulfilled only by zero-centered normal distributions $\mathcal{N}(0,\sigma)$.

A location parameter μ together with the previously defined three parameters α, β and γ fully characterize a stable distribution $p_{\alpha,\beta,\gamma,\mu}(x)$:

(i) *characteristic exponent*: $\alpha \in]\,0,2\,]$,
(ii) *skewness parameter*: $\beta \in [-1,1\,]$,
(iii) *scale parameter*: $\gamma \geq 0$, $\chi = \gamma^\alpha$, and
(iv) *location parameter*: $\mu \in \mathbb{R}$.

The parameters α and β determine the shape of the distribution and are called *shape parameters* therefore. The scale parameter γ is obviously the same as the standard deviation σ up to some factor, and the conventional mean μ serves as location parameter. The parameters of the three stable distributions with analytical densities are:

1. the normal distribution $\mathcal{N}(\mu,\sigma^2)$ with $\alpha = 2$, $\beta = 0$, $\gamma = \frac{\sigma}{\sqrt{2}}$, μ,
2. the Cauchy distribution $\mathcal{C}(\delta,\gamma)$ with $\alpha = 1$, $\beta = 0$, γ, μ and
3. the Lévy distribution $\mathcal{L}(\delta,\gamma)$ with $\alpha = \frac{1}{2}$, $\beta = 1$, γ, μ.

As we did in case of the normal distribution we define standard stable distributions having only two parameters by setting $\gamma = 1$ and $\mu = 0$:

$$p_{\alpha,\beta}(x) = p_{\alpha,\beta,1,0}(x) = p_{\alpha,\beta,1,0}\left(\frac{\xi-\mu}{\gamma}\right) = \gamma\, p_{\alpha,\beta,\gamma,\mu}(\xi)\ .$$

As said the characteristic exponent α is also called *index of stability* since it determines the order of the singularity at $u = 0$ and – at the same time – the long-distance scaling of the probability density [35, 66, 370]:

$$p_{\alpha,\beta}(x) \approx \frac{C(\alpha)}{|x|^{\alpha+1}} \quad \text{for} \quad x \to \pm\infty\ .$$

This scaling law is determined by the scaling parameter α, which turns out to be the spatial universality exponent ($\alpha = 2$ in case of the conventional diffusion in section 3.2.4).

Universality and self-similarity. Self-similarity and shapes of objects fitting fractal or non-integer dimensions are the major topics of Benoît Mandelbrot's

seminal book [303]. Self-similarity of stochastic processes has been mentioned already in section 3.2.4 in the context of continuous time random walks. In a nutshell looking at a self-similar object with a magnifying glass shows the same pattern no matter how large the magnification factor is. Needless to mention that real objects can be self-similar only over a few orders of magnitude, because resolutions cannot be increased without limits. The notion of *universality* was developed in statistical physics, in particular in the theory of phase transitions, where large collectives of atoms or molecules exhibit characteristic properties near critical points that are independent of the specific material parameters. Commonly power laws with critical exponents, $f(s) = f(s_0) \cdot |s - s_{\mathrm{crit}}|^{\alpha}$, are observed and when they are valid over several orders of magnitude the patterns become independent of the sizes of objects. Diffusion of molecules and condensation through aggregation are indeed examples of universal phenomena with the critical exponents $\alpha = 2$ in length and $\theta = 1$ in time. As said already the universality concerns the fact that all random walks with finite variance and finite waiting times fall into this universality class. With the experience gained from stable distributions and Lévy processes we can generalize the phenomena and compare the properties of processes with other universality exponents, $0 < \alpha \leq 2$ in space and $0 < \theta \leq 1$ in time. Higher exponents, $\alpha > 2$, are incompatible with normalizable probability densities. In particular convergence of the principal value integral cannot be achieved by a proper choice of $\delta(\epsilon)$ and ϵ [157, pp. 251-252].

The continuous time random walk (figure 3.16) is revisited here under the assumption of Lévy distributed jump lengths and waiting times. The calculation of the probability density $p\,(x,t)$ is in full analogy to that in section 3.2.4 and starts from the joint probability distribution $\varphi(\xi,\tau) = f(\xi)\psi(\tau)$ where independence according to (3.111) is assumed. The spatial increments are now derived from a stable Lévy distribution $f_{\alpha,0,\gamma,0}(\xi)$, which is symmetric ($\beta = 0$) and specified by the characteristic exponent, α, the scale parameter γ, and the location at the origin ($\mu = 0$). Since $f_{\alpha,0,\gamma,0}(\xi)$ need not be expressible in analytic form, we define it in terms of its characteristic function, which we write here in form of the limit of long distances or short k-values:[45]

$$\begin{aligned} \tilde{f}_{\alpha,\gamma}(|k|) \;&=\; \mathrm{E}\big(\exp\left(\mathrm{i}|k|\mathcal{X}_{\alpha,\gamma}\right)\big) \;=\; \int_{-\infty}^{\infty} \mathrm{d}\xi\, e^{\mathrm{i}|k|\xi} f_{\mathcal{X}_{\alpha,\gamma}}(\xi) \;= \\ &=\; \exp\left(-\gamma^{\alpha}|k|^{\alpha}\right) \;=\; 1 \,-\, \gamma^{\alpha}|k|^{\alpha} \,+\, \mathcal{O}(|k|^{2\alpha}) \;. \end{aligned} \tag{3.127}$$

The condition of obtaining an acceptable probability density – being nonnegative everywhere and normalizable – by inverse Fourier transform defines the range of possible values for the universality exponent: $0 < \alpha \leq 2$. We illustrate by means of examples and the obvious first example is the continuous

[45] The absolute value of the wave number $|k|$ is sometimes used in all expressions, which is necessary when complex k-values are admitted or in the multidimensional case where $\mathbf{k}$ is the wave vector. Here we use real k-values in one dimension and we need $|k|$ only to express a cusp at $k = 0$.

time random walk with $\alpha = 2$ and normally distributed jump length:

$$f(\xi) \;=\; \frac{e^{-\xi^2/4D}}{\sqrt{4\pi D}} \;\text{ and }\; \tilde{f}(k) \;=\; \exp\left(-\gamma^2 k^2\right) \;=\; 1 \,-\, \gamma^2 k^2 \,+\, \mathcal{O}(k^2)\;.$$

As a second illustrative example we mention the Cauchy distribution with the universality exponent $\alpha = 1$:

$$f(\xi) \;=\; \frac{\gamma}{\pi(\gamma^2 + \xi^2)} \;\text{ and }\; \tilde{f}(k) \;=\; \exp\left(-\gamma\,|k|\right) \;=\; 1 \,-\, \gamma\,|k| \,+\, \mathcal{O}(|k|^2)\;.$$

Since $-k^2$ decays of faster than $-|k|$, the Cauchy distribution – as shown already before – has heavier tails and sustains longer jumps.

The property of infinite divisibility can be used for a straightforward calculation of the mean length of a CTRW, $l = \overline{\mathcal{X}_0\mathcal{X}_n} = x_n = \sum_i \xi_i$, as expressed by means of the the *width* of the density $f_{\alpha,\gamma}(\frac{\xi}{n})$. Stability of the Lévy distribution requires that a linear combination of independent copies of the variable has the same distribution as the copy itself and this yields for the sum:

$$f_{n,\alpha,\gamma}\Big(\sum_i \xi_i\Big) \;=\; f_{\alpha,\gamma}(\xi_1) \circ f_{\alpha,\gamma}(\xi_2) \circ \ldots \circ f_{\alpha,\gamma}(\xi_n)\;,$$

where '$\circ$' stands for convolution. Application of the convolution theorem yields

$$\tilde{f}_n(|k|) \;=\; \prod_{i=1}^{n} \tilde{f}_{\alpha,\gamma}(|k_i|) \;=\; \exp\big(-\gamma\, n^{\frac{1}{\alpha}}\, |k|^{\alpha}\big)$$

Backtransformation into physical space yields a generalization of the central limit theorem:

$$f_{n,\alpha,\gamma}\Big(\sum_i x_i\Big) \;=\; f_{n,\alpha,\gamma}\Big(x \;/\; n^{\frac{1}{\alpha}}\Big)\;. \tag{3.128}$$

The length of the random walk is related to the width of the distribution, equation (3.128) yields the scaling of mean walk lengths: $l = \langle x(n)\rangle \propto n^{\frac{1}{\alpha}}$. In normal diffusion $\alpha = 2$ and the length grows with $\sqrt{n}$, for Lévy stable distributions with $\alpha < 2$ the walks become longer because of heavier tails compared to the normal distribution. The corresponding trajectories are called Lévy flights and will be discussed at the final paragraph of this section. In polymer theory the length of the walk corresponds to the end-to-end distance of the polymer chain for which analytic probability densities are available [414]. For polymers with Gaussian distributions, which follow from CLT for sufficiently long chains, the mean of the end-to-end distance fulfils a square root n law: $l \propto \sqrt{n}$.

The density of the waiting times is modified according the empirical evidence. There are well-documented processes with waiting times deviating from the expected exponential distribution in the sense that longer waiting times have higher probabilities or, in other words, the tails of the probability

densities decay slower than exponential. These deviations may have different origins, they are called *subdiffusion* and novel mathematical methods were developed in order to be able to deal with them properly [177, 328, 389]. In particular, adequate modeling of subdiffusion requires fractional calculus and since we shall not need this elegant but quite involved technique in this monograph, we dispense here from dwelling further on this discipline.

In order to take long rests or long-tails of the distribution of waiting times τ_w into account an asymptotic behavior of the form

$$\psi(\tau) \;\approx\; A_\theta \cdot \Big(\frac{\tau_w}{t}\Big)^{1+\theta} \quad \text{with} \;\; 0 < \theta \leq 1 \tag{3.129}$$

is assumed [328], which yields after Laplace transformation

$$\hat{\psi}(u) \;\approx\; \frac{1}{1+(\tau_w u)^\theta} \;=\; 1 \,-\, (\tau_w u)^\theta \,+\, \mathcal{O}(u^{1+\theta}) \tag{3.130}$$

The transformed joint distribution function can be obtained from the Montroll-Weiss equation (3.114):

$$\hat{\tilde{p}}\,(|k|,u) \;=\; \frac{1 \,-\, \hat{\psi}(u)}{u}\,\frac{1}{1 \,-\, \hat{\psi}(u)\,\tilde{f}(|k|)} \;\approx\; \frac{\tau_w\, u^{\theta-1}}{\tau_w u^\theta \,+\, \lambda|k|^\alpha}\;. \tag{3.131}$$

As we have seen in section 3.2.4 the expression on the r.h.s. of (3.131) with $\alpha = 2$, $\theta = 1$ can be subjected straightforwardly to inverse Laplace and Fourier transform, and this yields the density of normal diffusion $p\,(x,t) = \mathcal{L}^{-1}\mathcal{F}^{-1}\big(\hat{\tilde{p}}\,(|k|,u)\big) = e^{-x^2/4Dt}/\sqrt{4\pi Dt}$ (3.61).

For general Pareto processes the inverse Laplace and inverse Fourier transform on the expression of the r.h.s. of equation (3.131) is much more involved and cannot be completed in closed form. We can only indicate how one might proceed in the fractal case. For the inverse Laplace transform we get

$$\begin{aligned} p\,(x,t) \;&\approx\; \int_0^\infty \mathrm{d}u \int_{-\infty}^{+\infty} \mathrm{d}k\; e^{-\mathrm{i}|k|x+ut}\,\frac{\tau_w\, u^{\theta-1}}{\tau_w u^\theta \,+\, \lambda|k|^\alpha} \;= \\ &=\; \int_{-\infty}^{+\infty} \mathrm{d}k\; e^{-\mathrm{i}|k|x}\, E_\theta(-|k|^\alpha\, t^\theta)\;, \end{aligned} \tag{3.132}$$

where we made use of the Mittag-Leffler function $E_\theta(-|k|^\alpha\, t^\theta)$, which is named after Magnus Gösta Mittag-Leffter, occurs in inverse Laplace transforms of functions of the Laplace transform parameter $p^\alpha(a+bp^\beta)$ [311], and has the form of an infinite series [331]:

$$E_\alpha(z) \;=\; \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(1+\alpha k)}\,,\;\; \alpha \in \mathbb{C}, \Re(\alpha) > 0, z \in \mathbb{C}\;,$$

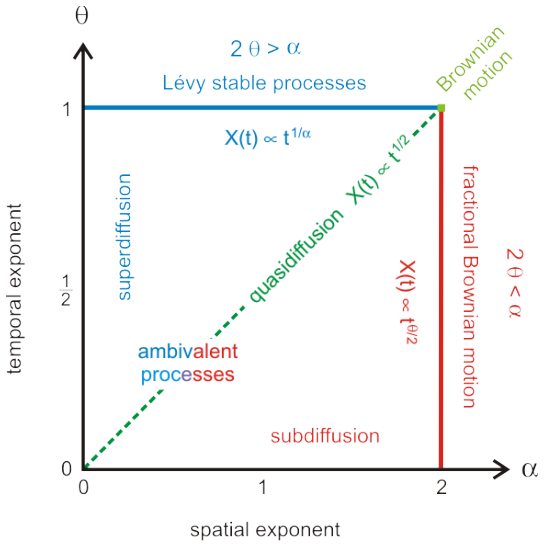


Fig. 3.19 Normal and anomalous diffusion. The figure sketches continuous time random walks (CTRW) as of the universality exponents of space, $0 < \alpha \leq 2$, and time, $0 < \theta \leq 1$. Lévy flights, normal diffusion, and fractional Brownian motion are limiting cases with the asymptotic behavior $(0 < \alpha < 2, \theta = 1)$, $(\alpha = 2, \theta = 1)$, and $(\alpha = 2, 0 < \theta < 1)$, respectively, of the general class of *ambivalent processes*.

which leads to quite involved expressions except in some simple cases, for example $E_1(z) = \exp(z)$ or $E_0(z) = 1/(1-z)$ [200]. The evaluation of the inverse Fourier transform (3.132) is even more complicated but we shall need to consider only the form of the leading terms: The function of the form $\tilde{p}\,(t^\theta |k|^\alpha)$in the integrand becomes a function $p\,(\frac{x^\alpha}{t^\theta})$ after the inverse Fourier transform. If we express distance as a function of time we obtain eventually: $\frac{x^\alpha}{t^\theta} = c \rightarrow x(t) \propto t^{\frac{\theta}{\alpha}}$. The expression covers normal diffusion with $\alpha = 2$ and $\theta = 1$ leading to the relation $x(t) \propto \sqrt{t}$ and fractional diffusion with $\alpha = 2$ and $\theta < 1$ resulting in $x(t) \propto t^{\theta/2}$.

In figure 3.19 we summarize the results of this section. All continuous time random walks are characterized by two universality exponents, $0 < \alpha \leq 2$ and $0 < \theta \leq 1$, for scaling behavior in space and time. Normal diffusion is the limiting case with $\alpha = 2$ and $\theta = 1$. The probability densities of time steps or waiting times and jump length, the Poisson distribution and the normal distribution, respectively, have both finite expectation values and variances. Lévy stable distributions with $\alpha < 2$ have heavy tails and the variance of the jump length diverges. Heavy tails makes larger jump increments more probable and the processes are characterized by longer walk lengths, $x(n) \propto n^{1/\alpha}$. Alternatively the variance of the step size is kept finite

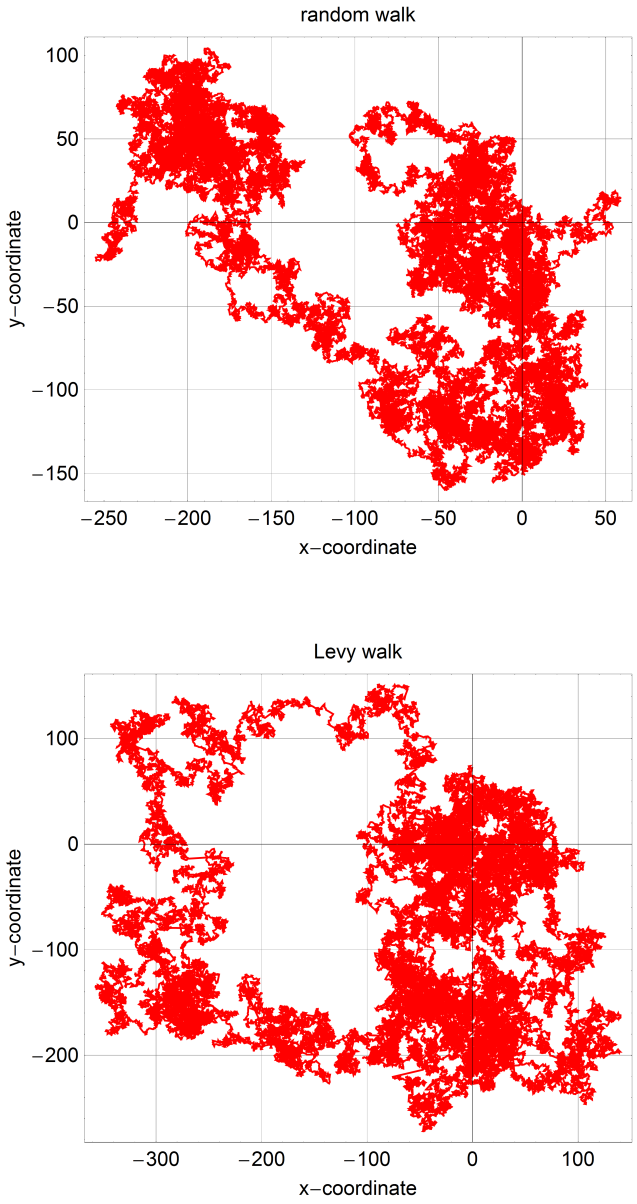


Fig. 3.20 Continued on next page.

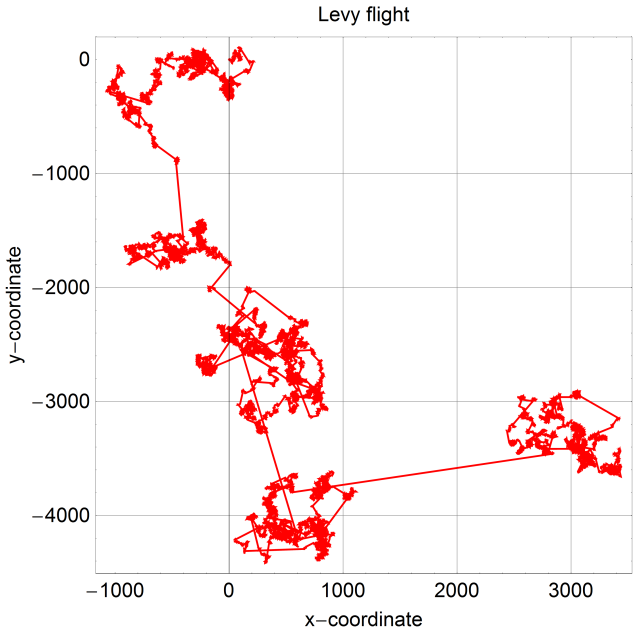


Fig. 3.20 Brownian motion and Lévy flights in two dimensions. The figure compares three trajectories of processes in the (x, y)-plane. Each trajectory consists of 100 000 incremental steps, which combines a direction that is randomly chosen from a uniform distribution: $\vartheta \in \mathcal{U}_\Omega$, $\Omega = [0, 2\pi]$ with a step length l. For the simulation of the random walk the step length was chosen to be $l = 1[l.u.]$ and for the Lévy flights the length was taken as a second set of random variables $l = \ell$, which were drawn from a density function $f_\ell(u) = u^{-(\alpha+1)}$ (3.122), and the components of the trajectory in the x- and y-direction are $x_{k+1} = x_k + l \cdot \cos\vartheta$ and $y_{k+1} = y_k + l \cdot \sin\vartheta$, respectively. The random variable ℓ is calculated from a uniformly distributed random variable v on $[0, 1]$ via the inverse cumulative distribution [89]:

$$\ell = F^{-1}(v) = u_m\,(1 - v)^{-1/(\alpha+1)}.$$

For a uniform density on $[0, 1]$ there is no difference in distribution between the random variables $(1 - v)$ and v and hence we used the simpler expression $\ell \propto v^{-1/(\alpha+1)}$ (The computation of pseudorandom numbers following a predefined distribution will be mentioned again in 4.6.3). The factor u_m is introduced as a lower bound for u in order to allow for normalization of the probability density, it can be interpreted as a scaling factor as well. Here we used $u_m = 1[l.u.]$. The examples shown were calculated with $\alpha = 2$ and $\alpha = 0.5$ and apostrophized by *Lévy walk* and *Lévy flight*, respectively. Apparently, there is no appreciable difference observable between the random walk and the Lévy walk. Random number generator: *Mersenne Twister* with the seeds '013' for random walk, '016' for Lévy walk, and '327' for Lévy flight.

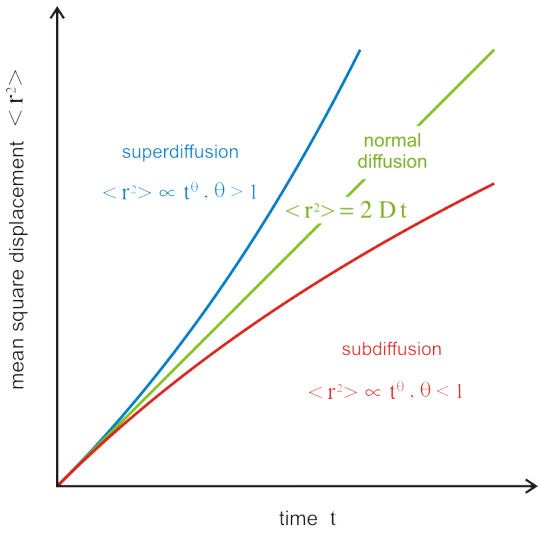


Fig. 3.21 Mean square displacement in normal and anomalous diffusion. The mean square displacement in normal diffusion is $\langle \mathbf{r}^2 \rangle = \langle (\mathbf{x} - \mathbf{x}_0)^2 \rangle = 2Dt$, and the generalization to anomalous diffusion allows for a classification of the processes according to the time exponent of the mean square displacement, $\langle \mathbf{r}^2 \rangle \propto t^\theta$: $\theta < 1$ characterizes *subdiffusion* and $\theta > 1$ *superdiffusion*.

in *anomalous diffusion* but the jumps are delayed and the waiting times diverge. The inner part of the square is filled by so-called *ambivalent processes* were the distributions of waiting times have diverging expectation values and no finite variances of the jump sizes (for details see [49, 328]).

Lévy processes derived from transition functions (3.122) with $0 < \alpha < 2$ correspond to densities with heavy tails and diverging variances. They were called *Lévy flights* by Benoît Mandelbrot [303]. Lévy flights with $\alpha = 2$ – called *Rayleigh flights* by Mandelbrot – turned out to be almost indistinguishable from conventional random walks with constant step size and accordingly both processes are suitable models for Brownian motion (figure 3.20). Since the Pareto transition function coincides with the normal, the Cauchy, and the Lévy distribution only in the asymptotic tails ($x \to \infty$) this similarity is a nice demonstration of the relevance of asymptotic behavior. In the limit $t \to \infty$, as already mentioned, 1D and 2D random walks lead to complete coverage of the line and the plane, respectively. Compared to the tails of the normal distribution the tails of all other Pareto transition functions, $\alpha < 2$, are heavier, and this implies higher probabilities for longer steps. In the special classes of Lévy flights with $\alpha = 1$ and $\alpha = 0.5$, for example, the step lengths may be randomly drawn from Cauchy or Lévy distributions, or derived from the power laws (3.122). The higher probabilities of long steps

changes completely the appearance of the trajectories: In the 2D plots densely visited zones are interrupted by occasional wide jumps that initiate a new local diffusion-like process in another part of the plane. In figure 3.20 we compare trajectories of 100 000 individual steps calculated by a random-walk routine with those computed with Levy flights with $\alpha = 2$ and $\alpha = 0.5$. The 2D pattern calculated for the Lévy flight with $\alpha = 2$ is very similar to the random walk pattern[46] whereas the Lévy flight with $\alpha = 0.5$ shows the expected small, more or less densely covered patches are separated by long jumps. It is illustrative to consider the physical dimensions of the area visited by the 2D-processes: The random walk and the Levy walk cover areas of approximately $300 \times 300\,[l.u.^2]$ and $400 \times 400\,[l.u.^2]$, but the Lévy flight ($\alpha = 0.5$) takes place in a much larger domain, $4000 \times 4000\,[l.u.^2]$.

The trajectories shown in figure 3.20 suggest to use the mean walk length for the classification of processes. From equation (3.128) followed a mean walk lengths $\langle x(n)\rangle \propto n^{1/\alpha}$ with n being the number of steps of the walk. Using the *mean square displacement* for the characterization of walk lengths we find for normal diffusion stating at the origin $\mathbf{x}(0) = \mathbf{x}_0 = 0$:

$$\left\langle \mathbf{r}(t)^2 \right\rangle_{\text{normal diffusion}} \;=\; \left\langle \big(\mathbf{x}(t) - \mathbf{x}_0\big)^2 \right\rangle \;=\; 2\,D\,t \;\propto\; t^{\theta} \;\text{ with }\; \theta = 1\;.$$

Anomalous diffusion is classified with respect to the asymptotic time dependence (figure 3.21): Subdiffusion is characterized by $\theta < 1$ and as mentioned above it is dealing with diffusion processes that are slowed down or delayed by structured environments. Superdiffusion with $\theta > 1$ is faster than normal diffusion and this is caused, for example, by a higher probability of longer jumps in a random walk.

The trajectory of the Lévy flight in figure 3.20 suggest an optimized search strategy: A certain area is searched thoroughly and after some while, for example when the territory has been harvested exhaustively, and then the search is continued in a rather distant zone. Prey foraging strategies of marine predators, for example those of sharks, were found to come close to Lévy flights. An optimal strategy consists in the combination of local searches by Brownian motion like movements and long jumps into distant regions where the next local search can start. The whole trajectory of such a combined search resembles the path of a Lévy flight [223, 448]. Finally, we repeat what we have mentioned initially: Lévy processes became an important issue in economics, in particular in finance [392].

[46] Because of this similarity we called the $\alpha = 2$ Pareto process as *Levy walk*.

3.3 Backward equations

Time inversion in a conventional differential equation changes the direction in which trajectories are passed through and this has only minor consequences for the phase portrait of the dynamical system: ω-limits become α-limits and vice versa, stable equilibrium points and limit cycles become unstable and so on, but the trajectories – without *the arrow of time* – remain unchanged. Apart from the arrow of time integrating forward yields precisely the same results as integrating backward from the endpoint of the forward trajectory. The same is true, of course, for a Liouville equation but it does not hold for a Wiener process or a Langevin equation: As sketched in figure 3.22 (lower part) time reversal results in trajectories that diverge in the backward direction. In other words, the commonly chosen reference conditions are such that a forward process has the sharp initial conditions at the beginning of the ordinary time scale – t_0 for t progressing into the future – whereas a backward process has sharp final conditions at the end – τ_0 for a virtual *computational time* τ progressing backwards into the past. Accordingly, the Chapman-Kolmogorov equation can be interpreted in two different ways giving rise to forward and backward equations that are equivalent to each other and the basic difference between both concerns the set of variables that is held fixed. In case on the forward equation we hold $(\mathbf{x}_0, t_0)$ fixed, and consequently solutions exist for $t \geq t_0$, so that $p\,(\mathbf{x}, t_0|\mathbf{x}_0, t_0) = \delta(\mathbf{x}-\mathbf{x}_0)$ is an *initial condition* for the forward equation. The backward equation on the other hand has solutions for $t \leq t_0$ corresponding $\tau \geq \tau_0$ and hence, it describes the evolution in τ. Accordingly, $p\,(\mathbf{y}, \tau_0|\mathbf{y}_0, \tau_0) = \delta(\mathbf{y} - \mathbf{y}_0)$ is an appropriate *final* condition (rather than an initial condition).[47]

Naïvely we could expect to find full symmetry between forward and backward computation, there is, however, one fundamental difference between calculations progressing in opposite directions, which will become evident when we consider backward equations in detail: In addition to the two different computational time scales for forward and backward equations – t and τ, respectively, in figure 3.22 – we have the real or physical time of the process, which has the same direction as t, unless we use some scaling factor it is even identical to t and we shall only distinguish the two time scales if necessary. The computational time τ, however, runs opposite to physical time, and the basic difference breaking the symmetry between forward and the backward equation thus concerns the arrow of time. The difference can also be expressed by saying the forward equations make prediction of the future and the backward equations reconstruct the past. In the eyes of mathematicians the backward equation is (somewhat) better defined than its forward analogue (see [131] and [134, pp. 321 ff.]).

[47] In order to avoid confusion we shall reserve the variable $y(\tau)$ and $y(0) = y_0$ for backward computation.

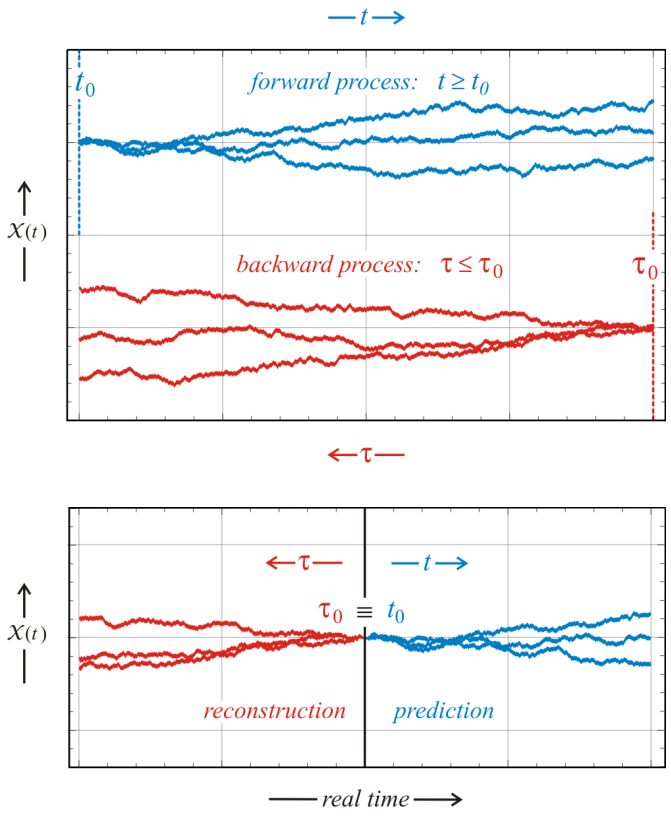


Fig. 3.22 Illustration of forward and backward equations. The forward differential Chapman-Kolmogorov equation is used in calculations of the future development of ensembles or populations. The trajectories (blue) start from an initial condition $(\mathbf{x}_0, t_0)$ commonly corresponding to the sharp distribution $p\,(\mathbf{x}, t_0) = \delta(\mathbf{x} - \mathbf{x}_0)$, and the probability density unfolds with time, $t \geq t_0$. The backward equation is commonly applied to the calculation of first passage times or the solution of exit problems. In order to minimize the risk of confusion we choose in backward equations the notation $\mathbf{y}$ and τ for the variable and the time, respectively, and we have the apparent correspondence $(\mathbf{y}(\tau), \tau) \Leftrightarrow (\mathbf{x}(t), t)$. In backward equations the latest time the corresponding value of the variable at this time, $(\mathbf{y}_0, \tau_0)$, are held constantτ_0 and a sharp initial condition – better called *final condition* in this case – is applied $p\,(\mathbf{y}, t_0 | \,\mathbf{y}, t) = \delta(\mathbf{y} - \mathbf{y}_0)$ and the time dependence of the probability density corresponds to samples unfolding into the past, $\tau \geq \tau_0$ (trajectories in red). In the lower part of the figure and alternative interpretation is given: The forward and the backward process start at the same time into different time directions, computation of the forward process makes predictions of the future whereas the backward process is calculated for the reconstruction of the past.

3.3.1 Backward Chapman-Kolmogorov equation

The Chapman-Kolmogorov equations (3.36 and 3.37) are interpreted in two different ways giving rise to the two formulations known as forward and backward equation. In the forward equation the double (x_3, t_3) is considered to be fixed and (x_1, t_1) expresses the variable in the sense of $x_1(t)$, where the time t_1 is progressing in the direction of positive real time (see figure 3.4). The backward equation, in contrary, is exploring the past of a given situation: Here, the double (x_1, t_1) is fixed and (x_3, t_3) is propagating backwards in time. The fact that real time proceeds in the forward direction has the consequence of somewhat different forms of forward and backward equations. Both Chapman-Kolmogorov differential expressions, the forward and the backward equation, are useful in their own rights. The forward equation gives directly the values of measurable quantities as functions of the *observed* or *real time*. Accordingly, it is preferentially used in describing actual processes and modeling experimental systems, and it is suited for predictions of probabilities in the future. The backward equation finds applications in the computation of the evolution towards given events, for example *first passage times* or *exit problems*, which are dealing with the search for the probability that a particle leaves a region at a certain time.

Since the difference in the derivation of forward and backward equations is essential for the interpretation of the results, we make a brief digression into the derivation of the backward equation, which is similar to but not identical with the procedure for the forward equation. The starting point again is the conditional probability of a Markov process from a recording $(\mathbf{y}, \tau)$ in the past to the final condition $(\mathbf{y}_0, \tau_0)$ at present: $p\,(\mathbf{y}_0, \tau_0 |\, \mathbf{y}, \tau_0) = \delta(\mathbf{y}_0 - \mathbf{y})$. As the term *backward* indicates we shall, however, assume that the computational time τ progresses from τ_0 into the past (figure 3.22) and the difference to the forward equation comes form the fact that computational and real time progress in opposite direction.

In order to avoid confusion we use real time in the derivation, proceed essentially in the same way as in section 3.2.1.2 and begin by writing down the infinitesimal limit of the difference equation:

$$
\begin{aligned}
&\frac{\partial p\,(\mathbf{y}_0, t_0 | \mathbf{y}, t)}{\partial t} \;= \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Big(p\,(\mathbf{y}_0, t_0 | \mathbf{y}, t{+}\Delta t) \;-\; p\,(\mathbf{y}_0, t_0 | \mathbf{y}, t) \Big) \;= \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_\Omega \mathrm{d}\mathbf{z}\; p\,(\mathbf{z}, t{+}\Delta t | \mathbf{y}, t) \Big(p\,(\mathbf{y}_0, t_0 | \mathbf{y}, t{+}\Delta t) \;-\; p\,(\mathbf{y}_0, t_0 | \mathbf{z}, t{+}\Delta t) \Big)
\end{aligned}
$$

where we have applied the same two operations as used for the derivation of equation (3.40): (i) resolution of unity,

$$
1 \;=\; \int_\Omega \mathrm{d}\mathbf{z}\, p\,(\mathbf{z}, t{+}\Delta t |\, \mathbf{y}, t) \;,
$$

and (ii) insertion of the Chapman-Kolmogorov equation in the second term with $\mathbf{z}$ being the intermediate variable,

$$p\,(\mathbf{y}_0, t_0|\mathbf{y}, t) \;=\; \int_\Omega \mathrm{d}\mathbf{z}\; p\,(\mathbf{y}_0, t_0|\mathbf{z}, t{+}\Delta t)\, p\,(\mathbf{z}, t{+}\Delta t|\mathbf{y}, t)\;.$$

Further steps parallel those in the derivation of the forward case: (i) separation of the domain of integration into two parts with the integrals I_1 and I_2 with $\|\mathbf{z}-\mathbf{y}\| < \epsilon$ and $\|\mathbf{z}-\mathbf{y}\| \geq \epsilon$, respectively, (ii) expansion of I_1 into a Taylor series, (iii) neglect of higher order residual terms, (iv) introduction of transition probabilities for jumps in the limit of vanishing Δt,

$$\lim_{\Delta t\to 0} \frac{1}{\Delta t}\; p\,(\mathbf{z}, t+\Delta t|\mathbf{y}, t) \;=\; W(\mathbf{z}|\mathbf{y}, t)\;. \tag{3.133}$$

(v) consideration of boundary effects if there are any, and eventually we obtain [157, pp. 55,56]:

$$\frac{\partial p\,(\mathbf{y}_0, t_0|\,\mathbf{y}, t)}{\partial t} \;=\; -\sum_i A_i(\mathbf{y}, t)\,\frac{\partial p\,(\mathbf{y}_0, t_0|\,\mathbf{y}, t)}{\partial y_i} + \tag{3.134a}$$

$$+\;\frac{1}{2}\sum_{i,j} B_{ij}(\mathbf{y}, \tau)\,\frac{\partial^2 p\,(\mathbf{y}_0, t_0|\,\mathbf{y}, t)}{\partial y_i \partial y_j} + \tag{3.134b}$$

$$+\int \mathrm{d}\mathbf{z}\; W(\mathbf{z}|\,\mathbf{y}, t)\,\Big(p\,(\mathbf{y}_0, t_0|\,\mathbf{y}, t) \;-\; p\,(\mathbf{y}_0, t_0|\,\mathbf{z}, t)\Big)\,. \tag{3.134c}$$

Equation (3.134) is called the *backward differential Chapman-Kolmogorov equation*, which complements the previously derived forward equation (3.46). The appropriate condition replacing equations (3.25) and (3.38) is

$$p\,(\mathbf{y}_0, t|\,\mathbf{y}, t) \;=\; \delta(\mathbf{y}_0 - \mathbf{y}) \quad \text{for all } t\,,$$

which expresses a sharp final conditions for $t = t_0$: $p\,(\mathbf{y}, t_0) = \delta(\mathbf{y}_0 - \mathbf{y})$ (figures 3.4 and 3.22). Apart from the change in sign of the drift term caused by $\Delta t = -\Delta\tau$, we realize changes in the structure of the PDE that make the equation in essence easier to handle than the forward equation. In particular, we find for the three terms:

The Liouville equation (section 3.2.2.1) is a partial differential equation whose physically relevant solutions coincide with the solution of an ordinary differential equation, and therefore the trajectories are invariant under time reversal – only the direction of the process is reversed: going backwards in time changes the signs of all components of $\mathbf{A}$ and the particle travels in opposite direction along the same trajectory that is determined by the initial or final condition $(\mathbf{x}_0, t_0)$ or $(\mathbf{y}_0, \tau_0)$.

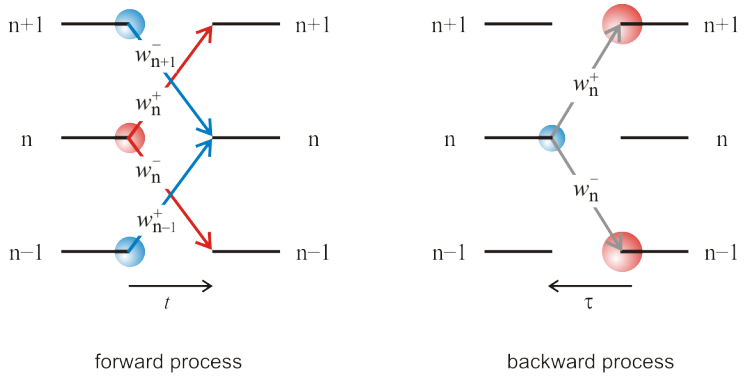


Fig. 3.23 Jumps in the single event master equations. The sketch on the left hand side shows the four single steps in the forward birth-and-death master equations, which are determined by the four transition probabilities w_n^+, w_{n-1}^+, w_{n+1}^-, and w_n^-. Transitions leading to a gain in probability P_n are indicated in blue, those reducing P_n are shown in red. On the right hand side we show the situation in the backward master equation: Only two transition probabilities, w_n^+ and w_n^- enter the equations, and the probabilities determining the amount of gain or loss in P_n are given at the final jump destinations rather than the beginnings.

The diffusion process described by equation (3.134b) spreads in opposite direction as a consequence of the inverse arrow of time. The mathematics of time reversal in diffusion has been studied extensively in the nineteen eighties [7, 110, 201, 408] and rigorous mathematical proofs were derived, which confirmed that inversion of time leads to indeed to a diffusion process in the direction of inverse time in the sense of the backward processes sketched in figure 3.22: Starting from a sharp final condition the trajectories diverge in the direction of $\tau = -t$.

The third term (3.134c) describes the jump processes and will be handled in the following section 3.3.2 on backward master equations..

3.3.2 Backward master equations

The backward master equation follows directly from the backward dCKE by setting $\mathbf{A} = 0$ and $\mathrm{B} = 0$ and this is tantamount to considering only the third term (3.134c). Since the difference in forward and backward equations is essential for the interpretation of the results, we consider the backward master equation in some detail. The starting point is the conditional probability of a Markov step process recorded from $(\mathbf{y}, \tau)$ in the past to the final condition $(\mathbf{y}_0, \tau_0)$ at present time τ_0: $p\,(\mathbf{y}_0, \tau_0 | \,\mathbf{y}, \tau_0) = \delta(\mathbf{y}_0 - \mathbf{y})$. As the term *backward* indicates we shall, however, assume that the computational time τ progresses

from τ_0 into the past. Some care is needed in applications to problem solution, because the direction of the time axis has influence on appearance and interpretation of transition probabilities. In computational time τ the jumps go in opposite direction (figure 3.23).

Equation (3.134c) in real time t yields on Riemann-Stieltjes integration:

$$\begin{aligned}\frac{\partial p\,(y_0,t_0|\,y,t)}{\partial t} &= \int_\Omega \mathrm{d}z\; W(z|\,y,t)\left(p\,(y_0,t_0|\,y,t) \;-\; p\,(y_0,t_0|\,z,t)\right) \;= \\ &= \sum_{z=0}^{\infty} W(z|\,y,t)\left(p\,(y_0,t_0|\,y,t) \;-\; p\,(y_0,t_0|\,z,t)\right) .\end{aligned}$$

Now we introduce the notation for discrete particle numbers, $y \Leftrightarrow n \in \mathbb{N}$, $z \Leftrightarrow m \in \mathbb{N}$, and $y_0 \Leftrightarrow n_0 \in \mathbb{N}$:

$$\frac{\mathrm{d}P_n(n_0,t_0|n,t)}{\mathrm{dt}} = \sum_{m=0}^{\infty} W(m|n,t)\left(P(n_0,t_0|n,t) \;-\; P(n_0,t_0|m,t)\right). \tag{3.135}$$

As previously we assume now time independent transition rates and restrict transitions to single births and deaths:

$$W(m|n,t) \;=\; W_{mn} \;=\; w_n^+\,\delta_{n+1,n} \;+\; w_n^-\,\delta_{n-1,n}\,, \quad \text{or}$$

$$W_{mn} \;=\; \begin{cases} w_n^+ & \text{if} \quad m = n+1\;, \\ w_n^- & \text{if} \quad m = n-1\;, \\ 0 & \text{otherwise}\;, \end{cases} \tag{3.95'}$$

Then, the backward single step master equation is of the form

$$\begin{aligned}\frac{\partial P(n_0,t_0|n,t)}{\partial t} &= w_n^+\Big(P(n_0,t_0|\,n,t) \;-\; P(n_0,t_0|\,n+1,t)\Big) \;+ \\ &\quad +\;\; w_n^-\Big(P(n_0,t_0|\,n,t) \;-\; P(n_0,t_0|\,n-1,t)\Big) \;= \\ &= -\,w_n^+\,P(n_0,t_0|\,n+1,t) \;-\; w_n^-\,P(n_0,t_0|\,n-1,t) \;+ \\ &\quad +\,(w_n^+ \;+\; w_n^-)\,P(n_0,t_0|\,n,t)\;.\end{aligned} \tag{3.136}$$

As in the case of the forward equation (3.97) the notation can be simplified by elimination of the common final state (n_0,t_0) and by making use of the discreteness of n:

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; w_n^+\Big(P_n(t) - P_{n+1}(t)\Big) \;+\; w_n^-\Big(P_n(t) - P_{n-1}(t)\Big) \tag{3.136'}$$

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; w_{n-1}^+\,P_{n-1}(t) \;+\; w_{n+1}^-\,P_{n+1}(t) \;-\; (w_n^+ + w_n^-)\,P_n(t)\;. \tag{3.97}$$

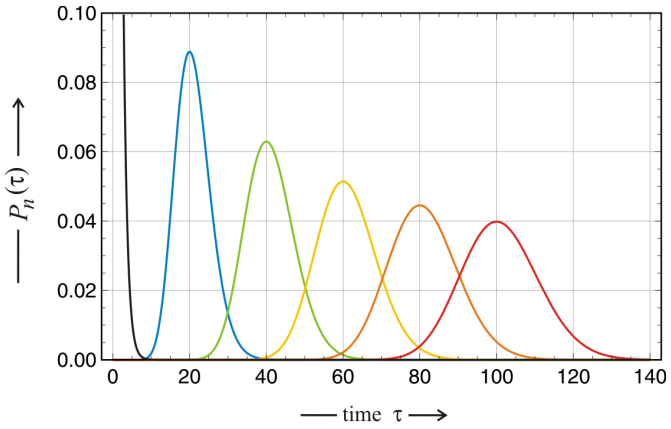


Fig. 3.24 Probability density of the backward Poisson process. The plot shows the probability density of the backward Poisson process, $P_n(\tau)$ (3.138), for different numbers of events $n = 0$ (black), 20 (blue), 40 (chartreuse), 60 (yellow), 80 (orange), and 100 (red). Further parameters: $n_0 = 100$ and $\gamma = 1$.

For the purpose of comparison we present the forward equation just below. The differences are visualized straightforwardly (figure 3.23). In the forward equation we need four transition probabilities, w_{n-1}^{+}, w_n^{-}, w_n^{+}, and w_{n+1}^{-} to describe the time derivative of the probability $P_n(t)$, and the interpretation of the terms for forward jumps is straightforward: The transition rates $w_k^{\pm}$ ($k = n-1, n, n+1$) are multiplied by the probabilities to be in the state before the jump at the instant of hopping. The calculation with the backward equation is simpler but the interpretation of the individual terms is more involved since the different directions of real time and computational time in the backward process change the situation: Only two transition probabilities appear in the backward equation and the probability terms are differences between the densities of two neighboring states, n and $n+1$ or n and $n-1$, respectively. The backward master equation is now applied to two different problems: (i) the backward Poisson process where we compute the first passage time of reaching the absorbing barrier of zero events, and (ii) a more general calculation of first passage times by means of the backward master equation.

3.3.3 Backward Poisson process

The relation between the solutions of the backward and forward master equations is illustrated for the Poisson process, which is sufficiently simple to be handled completely in closed form. The backward master equation of the

Poisson process is – as expected – closely related to the forward equation, since the transition probabilities are constant, $w_k^+ = w_k^- = \gamma \; \forall \; k$:

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; \gamma \Big(P_n(t) \;-\; P_{n+1}(t) \Big) \;\; \text{with} \;\; P_n(t_0) \;=\; \delta(n_0 - n) \; . \qquad (3.137)$$

Indeed, the forward equation is obtained in this simple case just by replacing $P_n(t)$ by $P_{n-1}(t)$ and $P_{n+1}(t)$ by $P_n(t)$. The backward Poisson process describes how n_0 events recorded at time t_0 could result from independent arrivals when the probability distribution of events follows an exponential density.

The solution of the master equation (3.137) is straightforward. Nevertheless, we repeat the technique based on probability generating function $g(s,t)$, because a few illustrative tricks are required. The expansion of $g(s,t)$ is not limited to positive n values [176, pp. 8-12]

$$g(s,t) \;=\; \sum_{n=-\infty}^{\infty} P_n(t)\, s^n \; ,$$

where the range of acceptable n values, $n_\mathrm{l} \leq n \leq n_\mathrm{h}, n \in \mathbb{Z}$ is introduced by setting $P_n(t) = 0 \, \forall \, n \notin [n_\mathrm{l}, n_\mathrm{h}]$. Insertion of equation (3.137), making use of the relation

$$\sum_{n=-\infty}^{\infty} P_{n+1}(t)\, s^n \;=\; \frac{1}{s} \sum_{n=-\infty}^{\infty} P_{n+1}(t)\, s^{n+1} \;=\; \frac{1}{s} \sum_{n=-\infty}^{\infty} P_n(t)\, s^n \; ,$$

and replacing $t_0 - t$ by the computational time τ yields the differential equation for the generating function that is a simple ODE because the expression does not contain derivatives with respect to the dummy variable s:

$$\frac{\mathrm{d}g(s,\tau)}{\mathrm{dt}} \;=\; \gamma \left(\frac{1}{s} - 1 \right) g(s,\tau) \quad \text{and} \quad g(s,t) \;=\; s^{n_0}\, e^{\gamma\tau/s}\, e^{-\gamma\tau} \; .$$

Taylor expansion of the second factor in $g(s,\tau)$ powers of $n_0 - n$ and equating coefficients yields the solution

$$P_n(\tau) \;=\; \frac{(\gamma\,\tau)^{n_0-n}}{(n_0-n)!}\, e^{-\gamma\tau} \; . \qquad (3.138)$$

In figure 3.24 we show the evolution of the probability density in the domain $0 \leq \tau \leq t_0$ corresponding to $t_0 \geq t \geq 0$. The computation of the moments of the random variable $\mathcal{N}(t)$ by means of equation 2.25 is a straightforward exercise in calculus and yields:

$$\mathrm{E}\big(\mathcal{N}(t)\big) \;=\; n_0 \;-\; \gamma\, t \;\; \text{and} \;\; var\big(\mathcal{N}(t)\big) \;=\; \gamma\, t \; . \qquad (3.139)$$

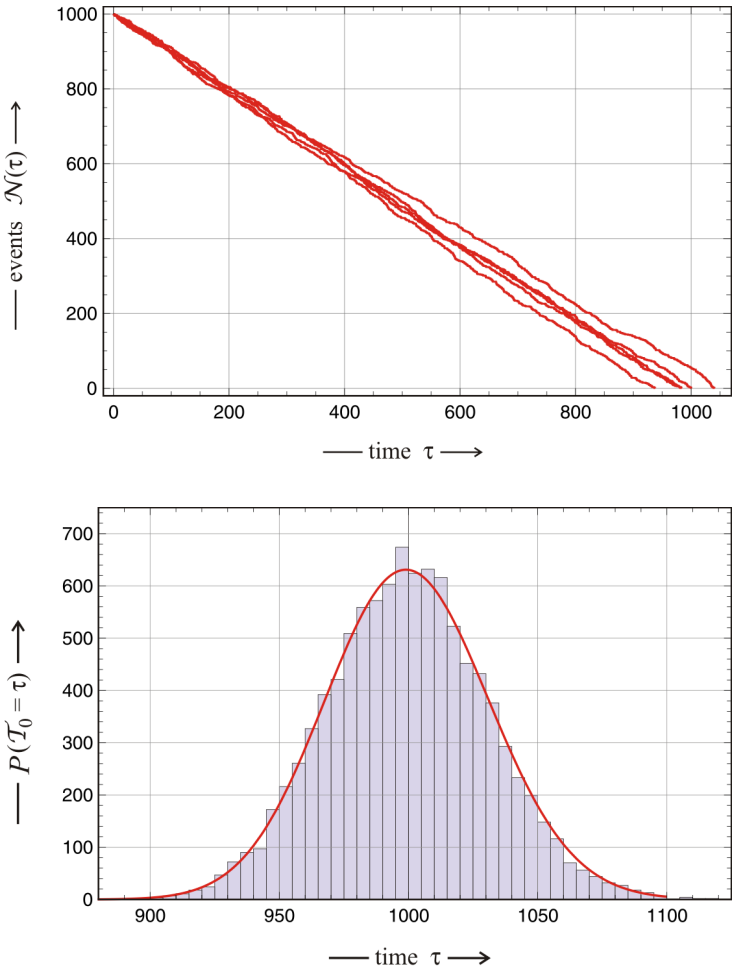


Fig. 3.25 The extinction time of the backward Poisson process. The upper part of the figure shows five trajectories of the backward Poisson process starting from $n_0 = 1000$ with $\gamma = 1$ and the seeds 013, 091, 491, 512, and 877, respectively. The lower part presents a histogram of the extinction times $\mathcal{T}_0$ obtained from 10 000 individual trajectories. The histogram is compared to the probability density of $\mathcal{T}_0$ which follows an Erlang distribution. Parameter values: $n_0 = 1000$ and $\gamma = 1$.

The linear time dependence is beautifully reflected by the trajectories shown in figure 3.25.

Next we make an attempt to calculate the time to reach $n = 0$ in a backward Poisson process. We shall call this kind of a first passage time the *extinction time*, $\mathcal{T}_0$ of the process in analogy to the notion of extinction times

in biology. The state with $n = 0$ represents an absorbing barrier: Once the process has reached this state it ends here, because the jumps in the backward Poisson are defined to fulfil $\Delta n = -1$. For the first passage times we adopt the notation of section 3.2.2.4 and find for the random variable

$$\mathcal{T}_0 \;=\; \sum_{k=1}^{n_0} \Delta t_k \;, \tag{3.140}$$

wherein Δt_k is the time span where exactly $n = k$ events were on the record. The probability of $\mathcal{T}_0$ lying between t and $t + \Delta t$ is given by the simultaneous occurrence of two events [435, pp. 71,72]: (i) One event is on the record, $P(\mathcal{N}(t) = 1)$, which implies that $n_0 - 1$ events have already taken place, and (ii) one further jump occurs, $P(\Delta\mathcal{N}(t) = -1)$

$$\begin{aligned} P(t \le \mathcal{T}_0 \le t + \Delta t) \;&=\; P\big(\mathcal{N}(t) = 1\big) \;\cdot\; P\big(\Delta\mathcal{N}(t) = -1\big) \;= \\ &=\; \frac{e^{-\gamma t}(\gamma\, t)^{n_0-1}}{(n_0-1)!} \;\cdot\; \gamma\, \Delta t \;= \\ &=\; \frac{\gamma^{n_0} t^{n_0-1}}{(n_0-1)!}\, e^{-\gamma\, t}\, \Delta t \;\; . \end{aligned}$$

Now we perform the limit $\Delta t \to \mathrm{dt}$ and find that the extinction time is distributed as

$$f_{\mathcal{T}_0} \;=\; \frac{\gamma^{n_0} t^{n_0-1}}{(n_0-1)!}\, e^{-\gamma\, t} \;, \tag{3.141}$$

which is known as Erlang distribution. It is straightforward to compute the expectation value and the variance of the extinction time

$$\mathrm{E}(\mathcal{T}_0) \;=\; \int_0^\infty t\, f_{\mathcal{T}_0}\, \mathrm{dt} \;=\; \frac{n_0}{\gamma} \;\; \text{and} \;\; \mathrm{var}(\mathcal{T}_0) \;=\; \frac{n_0}{\gamma^2} \;. \tag{3.142}$$

A comparison of numerically simulated extinction times and the analytical Erlang distribution is shown in figure 3.25. The numerical data meet the analytical expression as well as it could be. At the same time we see that the backwards process provides a natural access to the distribution of initial values, which lead to the final state (n_0, t_0).

3.3.4 Boundaries and mean first passage times

A *first passage time* is a random variable $\mathcal{T}$ that measures the instant when a particle passes a predefined location or state the first time and its expectation value $\mathrm{E}(\mathcal{T})$ is called *mean first passage time*. We need to stress *first*, because in a class of processes we are discussing here the variables may take on certain values finitely or in infinite time processes even infinitely often. A proper dis-

tinction can be made by considering boundaries for master equations, which fall into two classes: (i) *absorbing boundaries* and (ii) *reflecting boundaries*. When a particle or a process reaches an absorbing boundary it disappears or ends, respectively, whereas from a reflecting boundary it automatically returns to the domain of allowed values. Accordingly, an absorbing boundary can be reached only once whereas reflecting boundaries can be reached can be hit an infinite number of times. First passage times are not restricted to boundaries. Consider, for example, a random walk. As we pointed out every point on a straight line or a 2D-plane is visited an infinite number of times by any trajectory of infinite length.

Boundaries in birth-and-death master equations. The implementation of boundary conditions into single-step birth-and-death master equations is straightforward. The process is, for example, assumed to be restricted to the interval $a \leq n \leq b$, $n \in \mathbb{Z}$, and we only need to choose the appropriate transition probabilities that forbid the exit from the interval in case of a reflecting boundary or the return to the interval for an absorbing boundary (figure 3.26). Confining the process to $[a, b]$ we need two boundaries,[48] a lower boundary at $n = a$ and an upper one at $n = b$. Because of symmetry it is sufficient to consider only the lower boundary.

The boundary at $n = a$ is absorbing when the particle after it left the domain $[a, b]$ cannot return to it in forthcoming jumps, which is easily achieved by setting $w^+_{a-1} = 0$. A reflecting boundary results from the assumption $w^-_a = 0$: the particle cannot leave the domain. By symmetry we have $w^-_{b+1} = 0$ and $w^+_b = 0$ for the absorbing and the reflecting upper boundary.

In the forward single-step birth and death master equation the flux across the boundaries is only relevant for the equations of the states at the boundaries, $n = a$ and $n = b$. According to equation (3.97) with the initial condition $P_n(t_0) = \delta_{n,n_0}$ the differential change in the probability density at the lower boundary is

$$\frac{\mathrm{d}P_a(t)}{\mathrm{dt}} \;=\; w^-_{a+1}\,P_{a+1}(t) \,-\, w^+_a\,P_a(t) \,+\, w^+_{a-1}\,P_{a-1}(t) \,-\, w^-_a\,P_a(t)\;.$$

The two rightmost terms are effected by the boundary condition. In the case of reflection at the boundary the condition is that nothing flows out of the domain and this is fulfilled by $w^-_a = 0$ (figure 3.26), if the reflecting boundary is combined with no influx either w^+_{a-1} or $P_{a-1}(t)$ (or both) must be zero. In general the assumption $P_{a-1}(t) = 0$ is reasonable because it is not very meaningful to assume a finite probability density outside the domain $[a, b]$. Nevertheless, an influx can be modeled readily under the assumption of a virtual state $n = a - 1$. An alternative assumption is the equivalent to noflux or Neumann boundary conditions in partial differential equations: The flux

[48] Boundaries are also called *barriers* in the literature and both notions are used as synonyms. We shall use here exclusively the word 'boundary'. The expression barrier will be reserved for obstacles of motion inside the domain of the random variable.

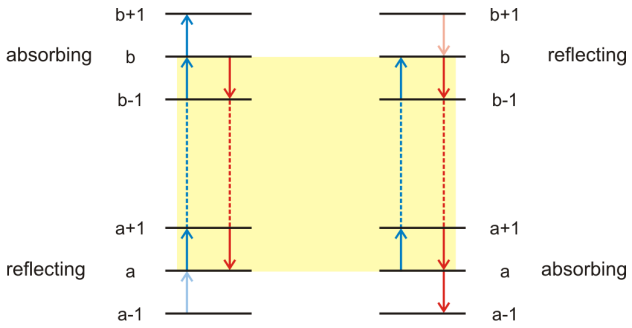


Fig. 3.26 Boundaries in single-step birth-and-death master equations. The figure on the l.h.s. sketches an interval, $a \leq n \leq b$ (indicated by yellow background), with a reflecting boundary at $n = a$ and an absorbing boundary at $n = b$ whereas the interval on the r.h.s. has the absorbing boundary at $n = a$ and the reflecting boundary at $n = b$. The step-up transition probabilities w_n^+ are shown in blue, the step-down transition probabilities w_n^- in red, a reflecting boundary has a zero outgoing probability, w_a^- or w_b^+, and the incoming probabilities, w_{a-1}^+ or w_{b+1}^-, are zero at an absorbing boundary. The incoming transition probabilities at the reflection boundaries are shown in light colors and play no role in the stochastic process because the probabilities of the corresponding virtual states are zero by definition: $P_{a-1}(t) = P_{b+1}(t) = 0$.

at the boundary has to vanish and this implies

$$w_{a-1}^+ \, P_{a-1}(t) \;=\; w_a^- \, P_a(t) \; . \tag{3.143}$$

Absorption at the lower boundary also allows for an alternative to setting $w_{a-1}^+ = 0$: Introducing a virtual state $n = a - 1$ and demanding $P_{a-1}(t) = 0$ yields the same effect. It is straightforward to show that the assumption of a virtual state $n = b + 1$ and the two conditions, $w_b^+ P_b(t) = w_{b+1}^- P_{b+1}(t) = 0$ and $P_{b+1} = 0$, do the same job for a reflecting or an absorbing upper barrier, respectively.

Alternative conditions can be found also for the backward master equation (3.136) on the interval $[a, b]$. At the lower boundary $n = a$ we find:

$$\begin{aligned} \frac{\mathrm{d}P(n_0, t_0|a, t)}{\mathrm{dt}} \;&= w_a^+ \, P(n_0, t_0|a+1, t) \;-\; w_a^+ \, P(n_0, t_0|a, t) + \\ &+ w_a^- \, P(n_0, t_0|a-1, t) \;-\; w_a^- \, P(n_0, t_0|a, t) \end{aligned}$$

for $n_0 \in [a, b]$. Again only the last two terms – the second line – are affected by the boundary conditions, and setting

$$P(n_0, t_0|a-1, t) \;=\; P(n_0, t_0|a, t) \tag{3.144}$$

is equivalent to putting $w_a^-(t) = 0$ in order to introduce a reflecting lower boundary through equating the second line to zero. The introduction of an absorbing lower boundary is a bit more tricky, since the transition rate w_{a-1}^+ does not appear in the backward master equation. Clearly, the condition $P(n_0, t_0|n, t) = 0$ with $n_0 \in [a, b]$ and $n < a$ will have the same effect as $w_{a-1}^+ = 0$. In single-step birth-and-death processes only the term with the largest value of n will be relevant for the process confined to the domain $[a, b]$ and hence $P(n_0, t_0|a-1, t) = 0$ is sufficient. At the upper boundary the corresponding two equations having the same effect as $w_b^+ = 0$ and $w_{b+1}^- = 0$ are: $P(n_0, t_0|b+1, t) = P(n_0, t_0|b, t)$ and $P(n_0, t_0|b+1, t) = 0$ for the reflecting and the absorbing boundary, respectively.

In this context it should be mentioned that in case of the chemical master equation equation we shall encounter *natural boundaries* where reaction kinetics itself takes care of reflecting or absorbing boundaries. If we are dealing with a reversible chemical reaction approaching a thermodynamic equilibrium in a system with a total number of N molecules the states $n_{\mathbf{K}} = 0$ and $n_{\mathbf{K}} = N$ are reflecting for each molecular species $\mathbf{K}$, whereas in an irreversible reaction the state $n_{\mathbf{K}} = 0$ is absorbing when the reactant $\mathbf{K}$ is at shortfall. Similarly, in absence of migration the state of extinction $n_{\mathbf{S}} = 0$ is an absorbing boundary for species $\mathbf{S}$.

First passage time in birth-and-death master equations. The calculation of a mean first passage time is illustrated by means of a simple example: The escape of a particle from a domain $[a, b]$ with a reflecting boundary at $n = a$ and an absorbing boundary at $n = b$ [301, pp. 90-92]. We make use of the backward master equation (3.136) and according to last paragraph we adopt the following conditions for the boundaries

$$P(n_0, t_0|a-1, t) \;=\; P(n_0, t_0|a, t) \;\text{ and }\; P(n_0, t_0|b+1, t) \;=\; 0 \;.$$

The probability that the particle is still in the interval $[a, b]$ is calculated by summation over all states in the accessible domain:

$$I_n(t) \;=\; \sum_{m=a}^{b} P(m, t|n, 0)\,,\; m \in \mathbb{Z}\;. \tag{3.145}$$

Insertion of the individual terms from the backward master equation (3.136) yields for the time derivative:

$$\begin{aligned} -\frac{\mathrm{d}I_n(t)}{\mathrm{dt}} \;&=\; \sum_{m=a}^{b} \frac{\mathrm{d}P(m, t|n, 0)}{\mathrm{dt}} \;= \\ &=\; w_n^+ \Big(I_n(t) \,-\, I_{n+1}(t)\Big) \,+\, w_n^- \Big(I_n(t) \,-\, I_{n-1}(t)\Big) \end{aligned} \tag{3.146}$$

with the conditions $I_{a-1}(t) = I_a(t)$ for the reflecting boundary at $n = a$ and $I_{b+1}(t) = 0$ for the absorbing boundary at $n = b$. The minus sign expresses the decrease in probability to be still within the interval $[a, b]$ in real time and is a consequence of the two time scales in backward processes, $\mathrm{dt} = -\,\mathrm{dt}$.

The probability of leaving the interval $[a, b]$ – the probability of absorption – within an infinitesimal interval of time $[t, t + \mathrm{dt}]$ is calculated to be

$$I_n(t) \;-\; I_n(t + \Delta t) \;=\; -\,\frac{\partial I_n}{\partial t}\;\mathrm{dt}\;,$$

and we can now obtain the mean first passage time for the escape from state n, $\langle \mathcal{T}_n \rangle$ by integration

$$\langle \mathcal{T}_n \rangle \;=\; -\int_0^\infty t\;\frac{\partial I_n}{\partial t}\;\mathrm{dt} \;=\; \int_0^\infty I_n\;\mathrm{dt}\;, \tag{3.147}$$

where the last expression results from integration by parts. Integration of equation (3.146) yields

$$\int_0^\infty -\,\frac{\partial I_n(t)}{\partial t}\,\mathrm{dt} \;=\; 1$$

for the l.h.s. since absorption of the particle or escape from the domain is certain. Integration of the r.h.s. yields mean passage times, and finally we obtain

$$1 \;=\; w_n^+ \left(\langle \mathcal{T}_n \rangle - \langle \mathcal{T}_{n+1} \rangle \right) \;+\; w_n^- \left(\langle \mathcal{T}_n \rangle - \langle \mathcal{T}_{n-1} \rangle \right) \tag{3.148}$$

the equation for the calculation of $\langle \mathcal{T}_n \rangle$. The boundary conditions are: $\langle \mathcal{T}_{a-1} \rangle = \langle \mathcal{T}_a \rangle$ and $\langle \mathcal{T}_{b+1} \rangle = 0$.

The solution of equation (3.148) for $\langle \mathcal{T}_n \rangle$ is facilitated by the introduction of new variables S_n and auxiliary functions φ_n:

$$S_n \;=\; \frac{\langle \mathcal{T}_{n+1} \rangle \;-\; \langle \mathcal{T}_n \rangle}{\varphi_n}\;,\; n \in [a, b] \;\text{ and}$$

$$\varphi_n \;=\; \prod_{m=a+1}^{n} \frac{w_m^-}{w_m^+}\;,\; n \in [a+1, b] \;\text{ with }\; \varphi_a \;=\; 1\;,$$

and in the new variables equation (3.148) takes on the form

$$-1 \;=\; w_n^+\,\phi_n \left(S_n \;-\; S_{n-1} \right)\;,$$

which allows for deriving a solution for the new variables

$$S_k \;=\; -\sum_{m=a}^{k} \frac{1}{w_m^+\,\varphi_m}\;.$$

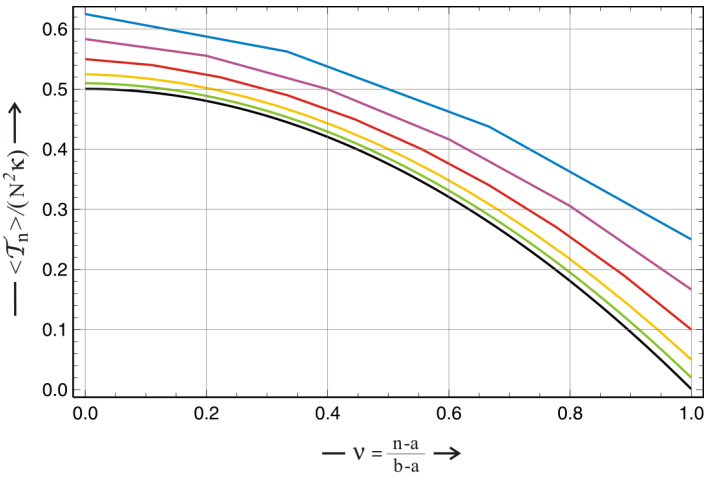


Fig. 3.27 Mean first passage times of a single-step birth-and-death process. Mean first passage times are computed from equation(3.150). In order to be able to compare the results for different sizes of the interval $[a, b]$ the interval is are normalized: $a = 0$ and $b = 1$, or $\nu = (n - a)/(b - a)$. Computed mean first passage times are scaled by a factor $(N^2\kappa)^{-1}$ with $N = b - a + 1$. The values for N chosen in the computations and the color code are: 4 (blue), 6 (violet), 10 (red), 20 (yellow), 50 (green), and 1000 (black).

From $\varphi_k S_k = \langle \mathcal{T}_{k+1} \rangle - \langle \mathcal{T}_k \rangle$ we obtain by means of the telescopic sum from $k = n$ to $k = b$

$$\sum_{k=n}^{b} \langle \mathcal{T}_{k+1} \rangle - \langle \mathcal{T}_k \rangle = \langle \mathcal{T}_{n+1} \rangle - \langle \mathcal{T}_n \rangle + \langle \mathcal{T}_{n+2} \rangle - \langle \mathcal{T}_{n+1} \rangle + \ldots + \langle \mathcal{T}_{b+1} \rangle - \langle \mathcal{T}_b \rangle =$$

$$= -\langle \mathcal{T}_n \rangle \ ,$$

because of the boundary condition $\langle \mathcal{T}_{b+1} \rangle = 0$, and we obtain the desired result

$$\langle \mathcal{T}_n \rangle \ = \ \sum_{k=n}^{b} \varphi_k \sum_{m=a}^{k} \frac{1}{w_m^+ \, \varphi_m} \ = \ \sum_{k=n}^{b} \frac{1}{w_k^+ \, \bar{P}_k} \sum_{m=a}^{k} \bar{P}_m \ , \tag{3.149}$$

where we have used the stationary probabilities $\bar{P}$ (3.100) instead of the functions φ to calculate the mean passage times.

For the purpose of illustration we choose a example that yields simple analytical expressions for the mean first passage times. The simplification is made with the transition probabilities:

$$w_k^+ \ = \ w_k^- \ = \ \kappa \ \ \forall \ \ k = 1, \ldots, N \ \ \text{and} \ \ w_0^+ = \kappa \, , \ w_0^- = 0 \, , \ w_{b+1}^- = 0 \ .$$

The number of states n in the interval $[a, b]$ with $a, b, n \in \mathbb{Z}$ is $N = b - a + 1$. Since $\bar{P}_n = \bar{P}_0$ follow from (3.100) we obtain by means of the normalization condition $\sum_n \bar{P}_n = 1$ the same probability $\bar{P}_n = 1/N \ \forall\, n$. Insertion in equation(3.149) yields the expression

$$\langle \mathcal{T}_n \rangle \ = \ \frac{1}{2\kappa} \, (b + n - 2a + 2)(b - n + 1) \ , \tag{3.150}$$

which has the leading term $-n^2$ in n. Numerical results are given in figure 3.27 and indeed the curves approach a negative quadratic function for large N.

Mean first passage times find widespread applications in chemistry and biology. Important study cases are the escape from potential traps, for example classical motion in the double well potential, fixation of alleles in population, and extinction times. We shall discuss examples in chapters 4 and 5.

3.4 Stochastic differential equations

The Chapman-Kolmogorov equation had been conceived in order to be able to model the propagation of probabilities in sample space. An alternative modeling approach (figure 3.1) starts out from a deterministic description by means of a difference or differential equation and superimposes a fluctuating random element. The idea of introducing stochasticity into deterministic modeling of processes goes back to the beginning of the twentieth century: In 1900 the French mathematician Louis Bachelier conceived and analyzed in his thesis the stochastic difference equation

$$\mathcal{X}(t_{n+1}) \;=\; \mathcal{X}(t_n) \,+\, \mu\,\Delta t \,+\, \sigma\sqrt{\Delta t}\; W_{n+1}$$

in order to model the fluctuating prices in the Paris stock exchange. Herein $\mu(\mathcal{X}_t, t)$ is a function related to the foreseeable development, $\sigma(\mathcal{X}_t, t)$ describes the amplitude of the random fluctuations and the W_n's are independent normal variables with mean zero and variance one in the sense of Brownian increments [24]. Remarkable is the fact that Bachelier's thesis preceded Einstein's and von Smoluchowski's famous works by five and six years, respectively.

The concept of stochastic differential equations is commonly attributed to the French mathematician Paul Langevin who proposed an equation named after him that allows for the introduction of random fluctuations into conventional differential equations [271]. The idea was to find a sufficiently simple approach to model Brownian motion successfully. In its original form the Langevin equation was written as

$$m\,\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{dt}^2} \;=\; -\,\gamma\,\frac{\mathrm{d}\mathbf{r}}{\mathrm{dt}} \,+\, \boldsymbol{\xi}(t) \quad \text{or} \quad \frac{\mathrm{d}\mathbf{p}\,(t)}{\mathrm{dt}} \;=\; -\frac{\gamma}{m}\,\mathbf{p}\,(t) \,+\, \boldsymbol{\xi}(t)\;. \qquad (3.151)$$

It describes the motion of a Brownian particle of mass m where $\mathbf{r}(t)$ and $\mathrm{d}\mathbf{r}/\,\mathrm{dt} = \mathbf{v}(t)$ are location and velocity of the particle, respectively. The term on the l.h.s. is the Newtonian gain in linear momentum $\mathbf{p}$ due to the force, $\mathrm{d}\mathbf{p}/\,\mathrm{dt}$, the first term on the r.h.s. is the loss of momentum due to friction, and the second term, $\boldsymbol{\xi}(t)$, represents the irregularly fluctuating *Brownian random force*. The Langevin equation can be written in terms of the momentum $\mathbf{p}$ and then it takes on the more familiar form. The parameter $\gamma = 6\pi\,\eta\,r$ is the friction coefficient according to Stokes law with η being the viscosity coefficient of the medium and r the size of the particle. The analogy of (3.151) to Newton's equation of motion is evident: The deterministic force, $f(x) = -(\partial V/\partial x)$ with $V(x)$ being the potential energy, is replaced by $\xi(t)$.

In figure 3.1 stochastic differential equations were shown as an alternative to the Chapman-Kolmogorov equation in modeling Markov processes. As said, the basic difference between the Chapman-Kolmogorov and the Langevin approach is the object whose time dependence is investigated: The Langevin equation 3.151 considers a single instant of a particle moving in

physical 3D-space that is exposed to thermal motion, and the integration yields a single stochastic trajectory. The Chapman-Kolmogorov equation of continuous motion leads to a Fokker-Planck equation (3.47), which describes the migration of a probability density in the same 3D-space where the trajectory is defined. *Equivalence* of both approaches expresses the fact that a sample of trajectories of a Langevin equation converges in distribution to the time dependent probability density of the corresponding Fokker-Planck equation in the limit of infinitely large samples. The equivalence of the Langevin and the Chapman-Kolmogorov approach is discussed in more detail in section 3.4.4. In case an analytical solution to a stochastic differential equation is available, the solution can be used to calculate moments of the probability distribution of $\mathcal{X}(t)$ and their time-dependence (section 3.4.5), especially mean and variance, which in practice are often sufficient for the description of a process.

In the literature one can find an enormous variety of detailed treatises of stochastic differential equations. We mention here the monograph [16] and two books that are available on the internet: [320, 352]. The forthcoming short sketch of stochastic differential equations follows in essence the line of thought chosen by Crispin Gardiner [157, pp.77-96].

3.4.1 Mathematics of stochastic differential equations

Generalization of equation (3.151) from Brownian motion to an arbitrary stochastic process yields

$$\frac{\mathrm{d}x}{\mathrm{dt}} \;=\; a(x,t) \,+\, b(x,t)\,\xi(t)\;, \tag{3.152}$$

where x is the variable under consideration and $\xi(t)$ is an irregularly fluctuating term often called *noise*. If the fluctuating term is independent of x, one speaks of *additive noise*. The two functions $a(x,t)$ and $b(x,t)$ are defined by the process to be investigated and the letters are chosen in order to point at an analogy to Fokker-Planck equations (3.47).

The theory of stochastic processes requires statistical independence for $\xi(t_1)$ and $\xi(t_2)$ if and only if $t_1 \neq t_2$. Furthermore we assume $\langle\xi(t)\rangle = 0$ without loosing generality since any drift term can be absorbed in $a(x,t)$, and encapsulate all requirements in an irregularity condition

$$\langle\xi(t_1)\,\xi(t_2)\rangle \;=\; \delta(t_1 - t_2)\;. \tag{3.153}$$

The Dirac δ-function diverges as $|t_1 - t_2| \to 0$ this has the consequence that $\langle\xi(t)\,\xi(t)\rangle$ and the variance $\mathrm{var}\big(\xi(t)\big) = \langle\xi(t)\,\xi(t)\rangle - \langle\xi(t)\rangle^2$ are infinite for $t_1 = t_2 = t$.

In order to be able to search for solutions of the differential equation (3.152) we require the existence of the integral

$$\omega(t) \;=\; \int_0^t \xi(t)\ \mathrm{dt}\ .$$

If $\omega(t)$ is a continuous function of time it has the Markov property, which can be proven by partitioning the integral

$$\begin{aligned}\omega(t_2) \;&=\; \omega(t_1) \,+\, \int_{t_1}^{t_2} \xi(\tau_2)\,\mathrm{d}\tau_2 \;=\; \int_0^{t_1} \xi(\tau_1)\,\mathrm{d}\tau_1 \,+\, \int_{t_1}^{t_2} \xi(\tau_2)\,\mathrm{d}\tau_2 \;= \\ &=\; \lim_{\varepsilon\to 0}\left(\int_0^{t_1-\varepsilon} \xi(\tau_1)\,\mathrm{d}\tau_1\right) \,+\, \int_{t_1}^{t_2} \xi(\tau_2)\,\mathrm{d}\tau_2\end{aligned}$$

and hence for every $\varepsilon > 0$ the $\xi(\tau_1)$ in the first integral is independent of the $\xi(\tau_2)$ in the second integral. By continuity $\omega(t_1)$ and $\omega(t_2) - \omega(t_1)$ are statistically independent in the limit $\varepsilon \to 0$, and further $\omega(t_2) - \omega(t_1)$ is independent of all $\omega(\vartheta)$ with $\vartheta < t_1$. In other words, $\omega(t_2)$ is completely determined in probabilistic terms by the value $\omega(t_1)$ and no information on past values is required: $\omega(t)$ is Markovian. □

From the experience gained with the derivation of the differential Chapman-Kolmogorov equation (3.46) and the postulated continuity of $\omega(t)$, we conjecture the existence a Fokker-Planck equation that describes $\omega(t)$. Computation of the drift and diffusion term is indeed straightforward [157, pp. 78,79] and yields $A(t) = 0$ and $B(t) = 1$:

$$\frac{\partial p\,(\omega,t)}{\partial t} \;=\; \frac{1}{2}\,\frac{\partial^2}{\partial\omega^2}\,p\,(\omega,t) \quad \text{with} \quad p\,(\omega,t_0) = \delta(\omega-\omega_0)\ . \tag{3.55}$$

Accordingly the Fokker-Planck equation describing the noise term in the Langevin equation is that of the Wiener process and we identify

$$\int_0^t \xi(\tau)\,\mathrm{d}\tau \;=\; \omega(t) \;=\; w(t)\ .$$

This innocent looking equation confronts us with a dilemma: As we know from the discussion of the Wiener process the solution of equation (3.55) $w(t)$ is continuous but *nowhere differentiable* and this has the consequence that neither $\mathrm{d}w(t)/\,\mathrm{dt}$ nor $\xi(t)$ nor the Langevin equation (3.151) nor the stochastic differential equation (3.152) exist in strict mathematical terms. Accessible to consistent interpretation in this rigorous sense is only the integral equation

$$x(t) \,-\, x(0) \;=\; \int_0^t a\big(x(\tau),\tau\big)\,\mathrm{d}\tau \,+\, \int_0^t b\big(x(\tau),\tau\big)\,\xi(\tau)\,\mathrm{d}\tau\ . \tag{3.154}$$

The relation to the Wiener process becomes more visible by writing

$$\mathrm{d}w(t) \;\equiv\; w(t+\mathrm{dt}) - w(t) \;=\; \xi(t)\,\mathrm{dt}\;,$$

and insertion into (3.154) yields the correct formulation:

$$x(t) - x(0) \;=\; \int_0^t a\big(x(\tau),\tau\big)\,\mathrm{d}\tau + \int_0^t b\big(x(\tau),\tau\big)\,\mathrm{d}w(\tau)\;. \tag{3.155}$$

The first integral is a conventional Riemann integral or a Riemann-Stieltjes integral if the function $a\big(x(\tau),\tau\big)$ contain steps, the second integral is a stochastic Stieltjes integral the evaluation of which will be discussed in the next section 3.4.2. In differential form we obtain for the formulation of the stochastic differential equation, which is compatible with standard mathematics:

$$\mathrm{d}x \;=\; a\big(x(t),t\big)\;\mathrm{dt} + b\big(x(t),t\big)\,\mathrm{d}w(t)\;. \tag{3.156}$$

The nature of the fluctuations is implicitly given by the differential Wiener process $\mathrm{d}w(t)$: The probability density is Gaussian corresponding to *white noise*. White noise is an idealization and provided more information on the physical background of the fluctuations is available $\mathrm{d}w(t)$ can be readily replaced by a more realistic noise term.

3.4.2 Stochastic integrals

A *stochastic integral* requires additional definitions compared to ordinary Riemann integration. We shall explain this rather unexpected fact and sketch some practical recipes for integration (for more details see [376]).

Definition of the stochastic integral. Let $G(t)$ be an arbitrary function of time and $w(t)$ the Wiener process, then the stochastic integral $I(t,t_0)$ is defined as a Riemann-Stieltjes integral (1.53) of the form

$$I(t,t_0) \;=\; \int_{t_0}^t G(\tau)\,\mathrm{d}w(\tau)\;. \tag{3.157}$$

The integral is partitioned into n subintervals, which are separated by the points t_i: $t_0 \le t_1 \le t_2 \le \cdots \le t_{n-1} \le t$ (figure 3.28). Intermediate points τ_i, which will be used for the evaluation of the function $G(\tau_i)$, are defined within each of the subintervals $t_{i-1} \le \tau_i \le t_i$ and – as it will be shown below – the value of the integral depends on the position chosen for τ_i within the subintervals.

The stochastic integral $\int_0^t G(\tau)\,\mathrm{d}w(\tau)$ is defined as the limit of the partial sums

$$S_n \;=\; \sum_{i=1}^{n} G(\tau_i)\,\Big(w(t_i) - w(t_{i-1})\Big)$$

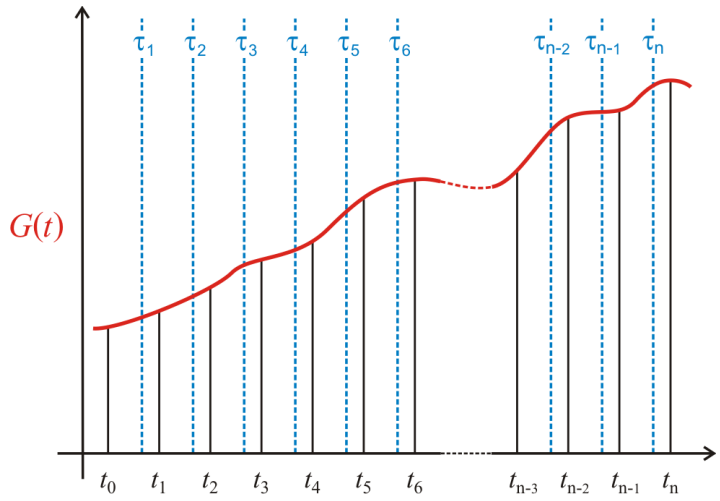


Fig. 3.28 Stochastic integral. The time interval $[t_0, t]$ is partitioned into n segments and an intermediate point τ_i is defined in each segment: $t_{i-1} \leq \tau_i \leq t_i$.

and it is not difficult to recognize that the integral is different for different choices of the intermediate point τ_i. As an important example we consider the case $G(t) = w(t)$:

$$\begin{aligned}
\langle S_n \rangle &= \left\langle \sum_{i=1}^{n} w(\tau_i) \Big(w(t_i) - w(t_{i-1}) \Big) \right\rangle = \\
&= \sum_{i=1}^{n} \langle w(\tau_i) w(t_i) \rangle - \sum_{i=1}^{n} \langle w(\tau_i) w(t_{i-1}) \rangle = \\
&= \sum_{i=1}^{n} \Big(\min(\tau_i, t_i) - \min(\tau_i, t_{i-1}) \Big) = \sum_{i=1}^{n} (\tau_i - t_{i-1}) \; .
\end{aligned}$$

As indicated in figure 3.28 we choose identical intermediate positions τ for all subintervals 'i'

$$\tau_i = \alpha\, t_i + (1-\alpha)\, t_{i-1} \quad \text{with} \quad 0 \leq \alpha \leq 1 \tag{3.158}$$

and obtain for the telescopic sum[49]

$$\langle S_n \rangle = \sum_{i=1}^{n} (t_i - t_{i-1})\, \alpha = (t - t_0)\, \alpha \; .$$

[49] In a telescopic sum all terms except the first and the last summand cancel.

Accordingly, the mean value of the integral may adopt any value between zero and $(t - t_0)$ depending on the choice of the position of the intermediate points as expressed by the parameter α. Out of possible choices two are popular leading to the Itō and the Stratonovich stochastic integral.

Itō stochastic integral. The most frequently used and most convenient definition of the stochastic integral is due to the Japanese mathematician Kiyoshi Itō [226, 227]. Semimartingales (section 3.1.3.2), in particular local martingales, are the most common stochastic processes that allow for straightforward application of Itō's formulation of stochastic calculus.

The choice $\alpha = 0$ or $\tau_i = t_{i-1}$ defines the Itō stochastic integral of a function $G(t)$:

$$\int_{t_0}^{t} G(\tau)\,\mathrm{d}w(\tau) \;=\; \lim_{n\to\infty} \sum_{i=1}^{n} G(t_{i-1})\Big(w(t_i) \,-\, w(t_{i-1})\Big)\;, \tag{3.159}$$

where the limit is a mean square limit (1.47). As an example we compute the previously discussed integral $\int_{t_0}^{t} w(\tau)\,\mathrm{d}w(\tau)$ and find for the sum S_n:

$$\begin{aligned} S_n \;&=\; \sum_{i=1}^{n} w(t_{i-1})\,\big(w(t_i) \,-\, w(t_{i-1})\big) \;\equiv\; \sum_{i=1}^{n} w(t_{i-1})\,\Delta w(t_i) \;= \\ &=\; \frac{1}{2}\sum_{i=1}^{n}\Big(\big(w(t_{i-1}) + \Delta w(t_i)\big)^2 \,-\, w(t_{i-1})^2 \,-\, \Delta w(t_i)^2\Big) \;= \\ &=\; \frac{1}{2}\Big(w(t)^2 \,-\, w(t_0)^2\Big) \,-\, \frac{1}{2}\sum_{i=1}^{n} \Delta w(t_i^2)\;, \end{aligned}$$

where the second line results from: $2\sum ab = \sum(a+b)^2 - \sum a^2 - \sum b^2$. It is now necessary to calculate the mean square limit of the second term in the last line of the equation. For a finite sum we have the expectation values

$$\left\langle \sum_{i=1}^{n} \Delta w(t_i)^2 \right\rangle \;=\; \sum_i \Big\langle \big(w(t_i) - w(t_{i-1})\big)^2 \Big\rangle \;=\; \sum_i (t_i - t_{i-1}) \;=\; t - t_0\;, \tag{3.160}$$

where the second equality results from the Gaussian nature of the probability density (3.61):[50]

$$\big\langle \big(w(t_i) - w(t_j)\big)^2 \big\rangle = \big\langle w(t_i)^2 \big\rangle - \big\langle w(t_j)^2 \big\rangle = \mathrm{var}\big(w(t_i)\big) - \mathrm{var}\big(w(t_j)\big) = t_i - t_j\;.$$

Next we calculate the expectation of the mean square deviation in (3.160):

[50] For the derivation of this relation we used the fact that the stochastic variables of the Wiener process at different times are uncorrelated, $\langle w(t_i)w(t_j)\rangle = 0$ and the variance is $\mathrm{var}\big(w(t_i)\big) = \langle w(t_i)^2\rangle - \langle w(t_i)\rangle^2 = \langle w(t_i)^2\rangle - \mu^2$.

$$\Big\langle \Big(\sum_{i=1}^{n} \big(w(t_i) - w(t_{i-1})\big)^2 \; - \; (t - t_0) \Big)^2 \Big\rangle \; =$$

$$= \Big\langle \sum_i \big(w(t_i) - w(t_{i-1})\big)^4 \; + \; 2 \sum_{i<j} \big(w(t_i) - w(t_{i-1})\big)^2 \big(w(t_j) - w(t_{j-1})\big)^2 \; -$$

$$- \; 2(t - t_0) \sum_i \big(w(t_i) - w(t_{i-1})\big)^2 \; + \; (t - t_0)^2 \Big\rangle \; .$$

We start the evaluation of the individual terms in the second line: According to (2.46) the fourth moment of a Gaussian variable can be expressed in terms of the variance

$$\Big\langle \big(w(t_i) - w(t_{i-1})\big)^4 \Big\rangle \; = \; 3 \Big\langle \big(w(t_i) - w(t_{i-1})\big)^2 \Big\rangle^2 \; = \; 3\,(t_i - t_{i-1})^2$$

Making use again of the independence of Gaussian variables we find

$$\Big\langle \big(w(t_i) - w(t_{i-1})\big)^2 \big(w(t_j) - w(t_{j-1})\big)^2 \Big\rangle \; = \; (t_i - t_{i-1})(t_j - t_{j-1}) \; .$$

Insertion into the expectation value eventually yields:

$$\Big\langle \Big(\sum_{i=1}^{n} \big(w(t_i) - w(t_{i-1})\big)^2 \; - \; (t - t_0) \Big)^2 \Big\rangle \; =$$

$$= \; 2 \sum_i (t_i - t_{i-1})^2 \; + \; \Big(\sum_i (t_i - t_{i-1}) - (t - t_0) \Big) \Big(\sum_j (t_j - t_{j-1}) - (t - t_0) \Big) \; =$$

$$= \; 2 \sum_i (t_i - t_{i-1})^2 \; \to \; 0 \;\; \text{as} \;\; n \to \infty \, ,$$

and, $\lim_{n\to\infty} \sum_i \big(w(t_i) - w(t_{i-1})\big)^2 = t - t_0$ in the mean square limit. □

The Itō stochastic integral of the Wiener process finally yields:

$$\int_{t_0}^{t} w(\tau) \, \mathrm{d}w(\tau) \; = \; \frac{1}{2} \Big(w(t)^2 - w(t_0)^2 - (t - t_0) \Big) \; . \tag{3.161}$$

Apparently the Itō integral differs from the conventional Riemann-Stieltjes integral where the term $t - t_0$ is absent. An illustrative explanation for this unusual behavior of the limit of the sum S_n is the fact that the quantity $|w(t + \Delta t) - w(t)|$ is almost always of the order $\sqrt{t}$ and hence – unlike in ordinary integration – the terms of second order in $\Delta w(t)$ do not vanish on taking the limit.

We remark that the expectation value of the integral (3.161) vanishes,

$$\Big\langle \int_{t_0}^{t} w(\tau) \, \mathrm{d}w(\tau) \Big\rangle \; = \; \frac{1}{2} \Big(\big\langle w(t)^2 \big\rangle - \big\langle w(t_0)^2 \big\rangle - (t - t_0) \Big) \; = \; 0 \; , \tag{3.162}$$

since the intermediate terms $\langle w(t_{i-1}) \Delta w(t_i) \rangle$ vanish because $\Delta w(t_i)$ and $w(t_{i-1})$ are statistically independent.

Nonanticipating functions. A stochastic process $\mathcal{X}(t)$ – as already mentioned in section 3.1.3.2 – is adapted or nonanticipating if and only if for every

trajectory and for every time t, $\mathcal{X}(t)$ is known at time t and *not before*.[51] In other words, a nonanticipating or adapted process *does not look into the future*, or a function $G(t)$ is nonanticipating or adapted to the process $\mathrm{d}w(t)$ if the value of $G(t)$ at time t depends only on the random increments $\mathrm{d}w(\tau)$ for $t \leq \tau$. Here we shall require this property in order to be able to solve certain classes of Itō stochastic integrals, which can be expressed as functions or functionals[52] of the Wiener process $w(t)$ by means of a stochastic differential or integral equation of the form

$$x(t) \,-\, x(t_0) \;=\; \int_{t_0}^{t} a\big(x(\tau),\tau\big)\,\mathrm{d}\tau \,+\, \int_{t_0}^{t} b\big(x(\tau),\tau\big)\,\mathrm{d}w(\tau)\;. \qquad (3.155)$$

A function $G(t)$ is *nonanticipating* with respect to t if $G(t)$ is probabilistically independent of $\big(w(s) - w(t)\big)$ for all s and t with $s > t$. In other words, $G(t)$ is independent of the behavior of the Wiener process in the future $s > t$. This is a natural and physically reasonable requirement for a solution of equation (3.155') because it boils down to the condition that $x(t)$ involves $w(\tau)$ only for $\tau \leq t$. Examples of important nonanticipating functions are

(i) $w(t)$,
(ii) $\int_{t_0}^{t} F\big(w(\tau)\big)\,\mathrm{d}\tau$,
(iii) $\int_{t_0}^{t} F\big(w(\tau)\big)\,\mathrm{d}w(\tau)$,
(iv) $\int_{t_0}^{t} G(\tau)\,\mathrm{d}\tau$, when $G(t)$ itself is nonanticipating, and
(v) $\int_{t_0}^{t} G(\tau)\,\mathrm{d}w(\tau)$, when $G(t)$ itself is nonanticipating.

The items (iii) and (v) depend on the fact that in Itō's version the stochastic integral is defined as the limit of a sequence in which $G(\tau)$ and $w(\tau)$ are involved exclusively for $\tau < t$.

Three reasons for the specific discussion of nonanticipating functions are important:
1. Results can be derived that are only valid for nonanticipating functions.
2. Nonanticipating functions occur naturally in situations, in which *causality* can be expected in the sense that the future cannot affect the presence.
3. The definition of stochastic differential equations requires nonanticipating functions.
In conventional calculus we never encounter situations in which the future acts back on the presence or even on the past.

Several relations are useful and required in Itō calculus:

[51] Every deterministic process is nonanticipating: In order to calculate the value $G(t + \mathrm{dt})$ of a function $t \to G(t)$ no value $G(\tau)$ with $\tau > t$ is required.

[52] A function assigns a value to the argument of the function, $x_0 \to f(x_0)$ whereas a functional relates a function to the value of a function, $f \to f(x_0)$.

$$\mathrm{d}w(t)^2 \;=\; \mathrm{d}t\,, \tag{3.163a}$$

$$\mathrm{d}w(t)^{2+n} \;=\; 0 \;\text{ for }\; n>0\,, \tag{3.163b}$$

$$\mathrm{d}w(t)\,\mathrm{d}t \;=\; 0\,, \tag{3.163c}$$

$$\int_{t_0}^{t} w(\tau)^n\,\mathrm{d}w(\tau) \;=\; \frac{1}{n+1}\Big(w(t)^{n+1}-w(t_0)^{n+1}\Big) - \frac{n}{2}\int_{t_0}^{t} w(\tau)^{n-1}\,\mathrm{d}\tau\,, \tag{3.163d}$$

$$\mathrm{d}f\big(w(t),t\big) \;=\; \left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial w^2}\right)\mathrm{d}t + \frac{\partial f}{\partial w}\,\mathrm{d}w(t)\;, \tag{3.163e}$$

$$\left\langle \int_{t_0}^{t} G(\tau)\,\mathrm{d}w(\tau)\right\rangle \;=\; 0\,,\quad \text{and} \tag{3.163f}$$

$$\left\langle \int_{t_0}^{t} G(\tau)\,\mathrm{d}w(\tau)\int_{t_0}^{t} H(\tau)\,\mathrm{d}w(\tau)\right\rangle \;=\; \int_{t_0}^{t}\langle G(\tau)H(\tau)\rangle\;\mathrm{d}\tau\;. \tag{3.163g}$$

The expressions are easier to memorize when we assign a dimension $[t^{1/2}]$ to $w(t)$ and discard all terms of order t^{1+n} with $n>0$.

Stratonovich stochastic integral. The value of a stochastic integral depends on the particular choice of the intermediate points, τ_i. The Russian physicist and engineer Ruslan Leontevich Stratonovich [417] and the American mathematician Donald LeRoy Fisk [144] developed simultaneously an alternative approach to Itō's stochastic integration, which is called Stratonovich integration:[53]

$$\fint_{t_0}^{t} G(\tau)\,\mathrm{d}w(\tau)$$

The intermediate points τ_i are chosen such that the unconventional term in the integral of $w(t)$, $(t-t_0)$ vanishes. For the purpose of illustration the integrand is chosen here again to be $G(t)=w(t)$, but now it is evaluated precisely in the middle of the interval, namely at time $\tau_i=(t_i-t_{i-1})/2$. Then, the mean square limit converges to the expression for the Stratonovich integral over $w(t)$

$$\begin{aligned}\fint_{t_0}^{t} w(\tau)\,\mathrm{d}w(\tau) \;&=\; \lim_{n\to\infty}\sum_{i=1}^{n}\frac{w(t_i)+w(t_{i-1})}{2}\Big(w(t_i)-w(t_{i-1})\Big) \;=\\ &=\; \frac{1}{2}\Big(w(t)^2-w(t_0)^2\Big)\;.\end{aligned} \tag{3.164}$$

[53] In order to distinguish the two versions of stochastic integrals we use the symbol $\int_{t_0}^{t}$ for the Itō integral and $\fint_{t_0}^{t}$ for the Stratonovich integral [231, p. 86]. The distinction from ordinary integrals is automatically provided by the differential $\mathrm{d}w$.

In contrast to the Itō integrals Stratonovich integration obeys the rules of conventional calculus but it is *not nonanticipating*, because the value of the function $w\big((t_{i-1}+t_i)/2\big)$ is already required at time t_{i-1}.

We compare here the derivation of both integral, Stratonovich and Itō [231, pp. 85-89], because additional insights are gained into the nature of stochastic processes. The starting point is the general Itō difference equation

$$\Delta x \;=\; F(x,t)\,\Delta t \,+\, G(x,t)\,\Delta w \;, \tag{3.165}$$

where $F(x,t)$ and $G(x,t)$ are functions defining drift and diffusion of the process under consideration, Δt and Δw are the time interval and the random increment, respectively. Next we choose equal time intervals as in figure 3.28 and have $t_k = k\,\Delta t + t_0$ with $x_k = x(t_k)$ and $\Delta x_{k-1} = x_k - x_{k-1}$ and $\Delta w_{k-1} = w_k - w_{k-1}$ where the starting point of the integration is: $t_0 = 0$, $x(0) = x_0$, $\Delta x_0 = x_1 - x_0$ and $\Delta w_0 = w_1 - w_0$ being the first random increment. Equation (3.165) takes on the precise form

$$\Delta x_{k-1} \;=\; F(x_{k-1},t_{k-1})\Delta t \,+\, G(x_{k-1},t_{k-1})\Delta w(t_{k-1})\,,\;\; k=1,\ldots,n\;,$$

Now we choose the starting point $t_0 = 0$ and $x(0) = x_0$, and find the general solution of the difference equation at $t = t_n$:

$$x(t_n) \;=\; x_n \;=\; x_0 \,+\, \sum_{k=0}^{n-1} F(x_k,t_k)\,\Delta t \,+\, \sum_{k=0}^{n-1} G(x_k,t_k)\,\Delta w(t_k)\;. \tag{3.166}$$

Equation (3.166) represents the explicit formula for the Cauchy-Euler integration (figure 3.29) and is commonly used in numerical SDE integration.

We use it here as basis for the explicit comparison of the Itō integral

$$\int_0^t G(x,t)\,\mathrm{d}w \;\equiv\; \lim_{n\to\infty} \sum_{k=0}^{n-1} G(x_k,t_k)\,\Delta w(t_k)\;, \tag{3.159'}$$

and the Stratonovich integral

$$\oint_0^t G\big(x,t\big)\,\mathrm{d}w \;\equiv\; \lim_{n\to\infty} \sum_{k=0}^{n-1} G\Big(\frac{x_{k+1}+x_k}{2},t_k\Big)\,\Delta w(t_k)\;, \tag{3.167}$$

and calculate the relationship between them. First we expand the function $G(x,t)$ in the Stratonovich analogue of the noise term in equation (3.166)

$$G\Big(\frac{x_{k+1}+x_k}{2},t_k\Big)\,\Delta w(t_k) \;=\; G\Big(x_k+\frac{\Delta x_k}{2},t_k\Big)\,\Delta w(t_k)\;,$$

in a power series around the point (x_n,t_n), and simplify the notation by defining $F_n \equiv F(x_n,t_n)$ and $G_n \equiv G(x_n,t_n)$ for the expansion,

$$G\Big(x_k + \frac{\Delta x_k}{2}, t_k\Big) \;=\; G_n \;+\; \frac{\partial G_n}{\partial x}\,\frac{\Delta x_n}{2} \;+\; \frac{1}{2}\,\frac{\partial^2 G_n}{\partial x^2}\,\Big(\frac{\Delta x_n}{2}\Big)^2 \;+\; \ldots \;.$$

Next we insert $\Delta x_n = F_n \Delta t + G_n \Delta w(t_n)$, recall that $\Delta w(t)^2 = \Delta t$, and find by omitting the higher order terms, because they will not contribute since all differences with higher powers, $(\Delta t)^\gamma$ with $\gamma > 1$ and $\big(\Delta w(t)\big)^\alpha$ with $\alpha > 2$ (3.163), vanish in the continuum limits $\Delta t \to \mathrm{dt}$ and $\Delta w \to \mathrm{d}w(t)$

$$G\Big(x_k + \frac{\Delta x_k}{2}, t_k\Big) \;=\; G_n + \left(\frac{F_n}{2}\frac{\partial G_n}{\partial x} + \frac{G_n^2}{4}\frac{\partial^2 G_n}{\partial x^2}\right)\Delta t + \frac{G_n}{2}\frac{\partial G_n}{\partial x}\,\Delta w(t_n)\,.$$

Next we insert this result into the discrete sum for the Stratonovich integral (3.167), omit the term with $\Delta t\,\Delta w$ since $\Delta t\,\Delta w \to \mathrm{dt}\,\mathrm{d}w(t) = 0$, and find

$$\sum_{k=0}^{n-1} G\Big(x_k + \frac{\Delta x_k}{2}, t_k\Big)\,\Delta w(t_k) \;=\; \sum_{k=0}^{n-1} G_k\,\Delta w(t_k) \;+\; \sum_{k=0}^{n-1} \frac{G_k}{2}\,\frac{\partial G_k}{\partial x}\,\Delta t\;.$$

Taking the continuum limit we obtain the desired relation between Itō and Stratonovich integrals

$$\oint_0^t G(x,t)\,\mathrm{d}w(t) \;=\; \int_0^t G(x,t)\,\mathrm{d}w(t) \;+\; \frac{1}{2}\int_0^t \frac{\partial G(x,t)}{\partial x}\,G(x,t)\;\mathrm{dt}\;. \quad (3.168)$$

The Stratonovich integral is equal to the Itō integral plus an additional contribution that can be assimilated into the drift term.

In summary we derived two integration methods for the stochastic differential equation

$$\mathrm{d}x \;=\; F(x,t)\;\mathrm{dt} \;+\; G(x,t)\,\mathrm{d}w(t) \;:\qquad (3.169)$$

(i) the Itō method yielding

$$x(t) \;=\; x(0) \;+\; \int_0^t F(x,t)\;\mathrm{dt} \;+\; \int_0^t G(x,t)\,\mathrm{d}w(t) \;\;\text{and}$$

(ii) the Stratonovich method resulting in a different solution, which we denote by $z(t)$ for the purpose of distinction

$$\begin{aligned} z(t) \;&=\; z(0) \;+\; \int_0^t F(z,t)\;\mathrm{dt} \;+\; \oint_0^t G(z,t)\,\mathrm{d}w(t) \;= \\ &=\; z(0) \;+\; \int_0^t \left(F(z,t) + \frac{G(z,t)}{2}\,\frac{\partial G(z,t)}{\partial z}\right)\;\mathrm{dt} \;+\; \int_0^t G(z,t)\,\mathrm{d}w(t)\;. \end{aligned}$$

On the other hand we would obtain the same solution $z(t)$ if we applied the Itō calculus to the stochastic differential equation

$$\mathrm{d}z \;=\; \left(F(z,t) + \frac{G(z,t)}{2}\,\frac{\partial G(z,t)}{\partial z} \right)\,\mathrm{dt} \;+\; G(z,t)\,\mathrm{d}w(t)\;. \tag{3.170}$$

Since the Stratonovich calculus is much more involved than the Itō calculus, we can readily see a strategy for obtaining Stratonovich solutions: Use equation (3.170) and derive the solution by means of Itō calculus. It is worth mentioning that a stand-alone Stratonovich integral has no relationship to a stand-alone Itō integral or, in other words, there is no connection between the two classes of integrals for an arbitrary function $G(t)$. Only when the stochastic differential equation is known to which the two integrals refer, a formula can be derived – as we did here – that relates the Itō integral to the Stratonovich integral.

At the end of this section we are left with the dilemma that the Itō integral is mathematically and technically most satisfactory but the more natural choice would be the Stratonovich integral that enables the usage of conventional calculus. In addition, the noise term in the Stratonovich interpretation can be real noise with finite correlation time whereas the idealized white noise required as reference in Itō's formalism gives rise to divergence of variances and correlations. The Stratonovich and not the Itō calculus, for example, is adequate for dealing with multiplicative noise in physical systems.

3.4.3 Integration of stochastic differential equations

A stochastic variable $x(t)$ is consistent with an Itō stochastic differential equation (SDE)

$$\mathrm{d}x(t) \;=\; a\big(x(t),t\big)\;\mathrm{dt} \;+\; b\big(x(t),t\big)\,\mathrm{d}w(t) \tag{3.156}$$

if for all t and t_0 the integral equation (3.155) is fulfilled. Time is ordered,

$$t_0 \;<\; t_1 \;<\; t_2 \;<\; \cdots \;<\; t_n \;=\; t\;,$$

and the time axis may be assumed to be split into – equal or unequal – increments, $\Delta t_i = t_{i+1} - t_i$. We visualize a particular solution curve of the SDE for the initial condition $x(t_0) = x_0$ in the spirit of a discretized version

$$x_{i+1} \;=\; x_i \;+\; a(x_i,t_i)\,\Delta t_i \;+\; b(x_i,t_i)\,\Delta w(t_i)\;, \tag{3.155'}$$

wherein $x_i = x(t_i)$, $\Delta t_i = t_{i+1} - t_i$, and $\Delta w(t_i) = w(t_{i+1}) - w(t_i)$. Figure 3.29 illustrates the partitioning of the stochastic process into a deterministic drift component, which is the discretized solution curve of the ODE obtained by setting $b\big(x(t),t\big) = 0$ in equation (3.155'), and a stochastic diffusion compo-

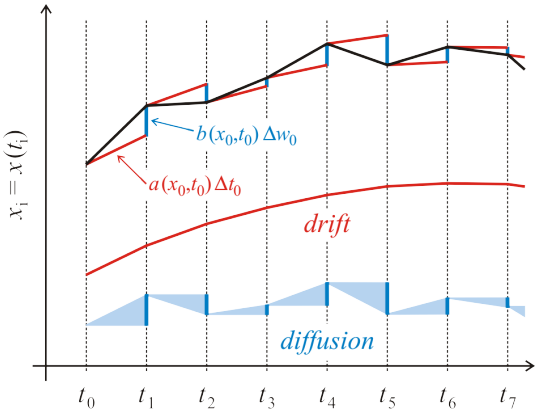


Fig. 3.29 Stochastic integration. The figure sketches the Cauchy-Euler procedure for the construction of an approximate solution of the stochastic differential equation (3.152'). The stochastic process consists of two different components: (i) the drift term, which is the solution of the ODE in absence of diffusion (red; $b(x_i, t_i) = 0$) and (ii) the diffusion term representing a Wiener process $w(t)$ (blue; $a(x_i, t_i) = 0$). The superposition of the two terms gives the stochastic process (black). The two lower plots show the two components in separation. The increments of the Wiener process $\Delta w(t_i)$ are uncorrelated and independent. An approximation to a particular solution of the stochastic process is constructed by letting the mesh size approach zero, $\lim \Delta t \to 0$.

nent, which is a Wiener process $w(t)$ that is obtained by setting $a\big(x(t), t\big) = 0$ in the SDE. The increment of the Wiener process, $\Delta w(t_i)$, is independent of x_i provided (i) x_0 is independent of all $w(t) - w(t_0)$ for $t > t_0$ and (ii) $a(x, t)$ is a nonanticipating function of t for any fixed x. Condition (i) is tantamount to the requirement that any random initial condition must be nonanticipating.

In the construction of approximate solutions $x(t)$ in discretized form the value $x_i = x(t_i)$ is always independent of $\Delta w(t_j)$ for $j \geq i$ as we verify easily by inspection of (3.155') or considering figure 3.29:

$$x_i \;=\; x_{i-1} \,+\, a(x_{i-1}, t_{i-1})\,\Delta t_{i-1} \,+\, b(x_{i-1}, t_{i-1})\,\Delta w(t_{i-1}) \;.$$

A particular solution to equation (3.156) is derived through performing the limit of vanishing mesh size: $\lim \Delta t \to 0$, which implies $\lim n \to \infty$. Uniqueness of solutions refers to individual trajectories in the sense that a particular solution is uniquely obtained for a given sample function $\mathcal{W}_k(t)$ of the Wiener Process $w(t)$. The existence of a solution, however, is defined for the whole ensemble of sample functions: A solution of equation (3.156) exists if a particular solution exists with probability one for any choice of a sample function of the Wiener process $\mathcal{W}_k(t)\,\forall\, k \in \mathbb{N}_{>0}$. Existence and uniqueness of a nonan-

ticipating solution $x(t)$ of an Itō stochastic differential equations within the time interval $[t_0, t]$ can be proven for two conditions [16, pp.100-115]:

(i) *Lipschitz condition*: there exists a Lipschitz constant $L > 0$ such that

$$|a(x,\tau) - a(y,t)| \; + \; |b(x,t) - b(y,t)| \; \leq \; L\,|x - y|$$

for all x and y and $t \in [t_0, t]$, and

(ii) *linear growth condition*: a κ exists such that for all $t \in [t_0, t]$

$$|a(x,t)|^2 \; + \; |b(x,t)|^2 \; \leq \; \kappa^2\,(1 + |x|^2) \; .$$

The *Lipschitz condition* named after the German mathematician Rudolf Lipschitz ensures existence and uniqueness of the solution and is almost always fulfilled for stochastic differential equations in practice, because in essence it is a smoothness condition. The *linear growth condition* guarantees boundedness of the solution (For details see, for example, [352, pp. 68-71]. The growth condition may be violated in abstract model equations, for example, when a solution *explodes* and progresses to infinity at finite time. A very simplex example is given by

$$\frac{\mathrm{d}x}{\mathrm{dt}} \; = \; x^2 \;\; \text{with} \;\; x(0) \, = \, 1 \;\; \Longrightarrow \;\; x(t) \, = \, \frac{1}{1-t} \;\; \text{for} \;\; 0 \leq t < 1 \; ,$$

which is unbounded at $t = 1$ and has no global solution defined for all values of t (see section 5.1.3) or in other words, the value of x becomes infinite at some finite time. We shall encounter such situations in chapter 5. As a matter of fact this is a typical model behavior since no population or spatial variable can approach infinity at finite times in a finite world.

Several other properties known to apply to solutions of ordinary differential equations can be shown without major modifications to apply to SDE's too: Continuity in the dependence on parameters and boundary conditions as well as the Markov property (for proofs we refer to [16]).

3.4.4 Some properties of stochastic calculus

Changing variables is a technical issue but important for applications and boring when one makes errors. Since Itō calculus is different from ordinary calculus, we expect differences also in the rules of substituting variables. In order to see the general effect of substitutions in Itō's stochastic differential equations we consider an arbitrary function, $\mathbf{x}(t) \Rightarrow f\big(\mathbf{x}(t)\big)$, and calculate $\mathrm{d}x(t) \Rightarrow \mathrm{d}f\big(\mathbf{x}(t)\big)$. The major difference compared to ordinary calculus comes from the necessity to extend all expansions up to second order because $\mathrm{d}w(t)^2 \, = \, \mathrm{dt}$ and hence $\Delta w(t)^2$ does not approach zero faster than Δt in

the limit $\Delta t \to \mathrm{dt}$. We start with the simpler case of a single variable and afterwards introduce the multidimensional situation.

Change of variables. Starting out from the SDE $\mathrm{d}x = a(x,t)\,\mathrm{dt} + b(x,t)\mathrm{d}w(t)$ and making use of our previous results on nonanticipating functions we expand $\mathrm{d}f\big(x(t)\big)$ up to second order but retain only the term in $\mathrm{d}w(t)$, because by the Itō rules we have $\mathrm{dt}^2 = 0$ and $\mathrm{d}w(t)\,\mathrm{dt} = 0$ (and write x instead $x(t)$):

$$\begin{aligned} \mathrm{d}f(x) \;&=\; f(x+\mathrm{d}x) \,-\, f(x) \;= \\ &=\; \frac{\partial f(x)}{\partial x}\,\mathrm{d}x \,+\, \frac{1}{2}\frac{\partial^2 f(x)}{\partial x^2}\,\mathrm{d}x^2 \,+\, \cdots \;= \\ &=\; \frac{\partial f(x)}{\partial x}\Big(a(x,t)\,\mathrm{dt} \,+\, b(x,t)\,\mathrm{d}w(t)\Big) \,+\, \frac{1}{2}\frac{\partial^2 f(x)}{\partial x^2}\,b(x,t)^2\,\mathrm{d}w(t)^2\,, \end{aligned} \tag{3.171}$$

where all terms higher than second order have been neglected. According to Itō calculus (3.163) we introduce $\mathrm{d}w(t)^2 = \mathrm{d}t$ into the last line of this equation and obtain Itō's formula:

$$\begin{aligned} \mathrm{d}f\big(x(t)\big) \;=\; &\Big(a\big(x(t),t\big)\,\frac{\partial f\big(x(t)\big)}{\partial x} \,+\, \frac{1}{2}\,b\big(x(t),t\big)^2\,\frac{\partial^2 f\big(x(t)\big)}{\partial x^2}\Big)\,\mathrm{d}t \,+ \\ &+\, b\big(x(t),t\big)\,\frac{\partial\big(x(t)\big)}{\partial x}\,\mathrm{d}w(t)\;. \end{aligned} \tag{3.172}$$

It is worth noticing that Itō's formula and ordinary calculus lead to different results unless $f(x)$ is linear in $x(t)$ and accordingly $\frac{\partial^2 f(x)}{\partial x^2}$ vanishes.

As an exercise we suggest to calculate the substitution by the function $f(x) = x^2$. The result is

$$\mathrm{d}(x^2) \;=\; \big(2x\,a(x,t) \,+\, b(x,t)^2\big)\;\mathrm{dt} \,+\, 2\,b(x,t)\,\mathrm{d}w(t)\;,$$

which is, for example, useful to calculate the time derivative of the variance: $\mathrm{d}\,\mathrm{var}\big(x(t)\big)/\,\mathrm{dt} = \mathrm{d}\,\big\langle x^2\big\rangle/\,\mathrm{dt} + 2\,\langle x\rangle\;\mathrm{d}\,\langle x\rangle/\,\mathrm{dt}$.

The application of Itō's formalism to many dimensions, in general, becomes very complicated. The most straightforward simplification is the extension of Itō calculus to the multivariate case by making use of the rule that $\mathrm{d}w(t)$ is an infinitesimal of order $t^{1/2}$. Then we can show that the following relations hold for an n-dimensional Wiener process $\mathbf{w}(t) = \big(w_1(t), w_2(t), \ldots, w_n(t)\big)$:

$$\mathrm{d}w_i(t)\,\mathrm{d}w_j(t) \;=\; \delta_{ij}\;\mathrm{dt}\;, \tag{3.173a}$$

$$\mathrm{d}w_i(t)^{2+N} \;=\; 0\;, \qquad (N>0)\;, \tag{3.173b}$$

$$\mathrm{d}w_i(t)\;\mathrm{dt} \;=\; 0\;, \tag{3.173c}$$

$$\mathrm{dt}^{1+N} \;=\; 0\;, \qquad (N>0)\;. \tag{3.173d}$$

The first relation is a consequence of the independence of increments of Wiener processes along different coordinate axes, $\mathrm{d}w_i(t)$ and $\mathrm{d}w_j(t)$. Making

use of the drift vector $\mathbf{A}(\mathbf{x},t)$ and the diffusion matrix $\mathrm{B}(\mathbf{x},t)$ the multidimensional stochastic differential equation

$$\mathrm{d}\mathbf{x} \;=\; \mathbf{A}(\mathbf{x},t)\;\mathrm{dt} \;+\; \mathrm{B}(\mathbf{x},t)\,\mathrm{d}\mathbf{w}(t)\;. \tag{3.174}$$

Following Itō's procedure we obtain for an arbitrary well-behaved function $f\big(\mathbf{x}(t)\big)$ the result

$$\begin{aligned}\mathrm{d}f(\mathbf{x}) \;=\; &\Bigg(\sum_i A_i(x,t)\,\frac{\partial}{\partial x_i}f(\mathbf{x}) + \\ &+\;\frac{1}{2}\sum_{i,j}\big(\mathrm{B}(x,t)\cdot\mathrm{B}'(x,t)\big)_{ij}\;\frac{\partial^2}{\partial x_i\partial x_j}f(\mathbf{x})\Bigg)\,\mathrm{d}t + \\ &+\;\sum_{i,j}B_{ij}\frac{\partial}{\partial x_i}f(\mathbf{x})\,\mathrm{d}w_j(t)\;.\end{aligned} \tag{3.175}$$

Again we observe the additional term introduced through the definition of the Itō integral.

Fokker-Planck equations and SDEs. The expectation value of an arbitrary function $f\big(x(t)\big)$ can be calculated by means of Itō's formula. We begin with a single variable:

$$\begin{aligned}\frac{\big\langle \mathrm{d}f\big(x(t)\big)\big\rangle}{\mathrm{d}t} \;&=\; \left\langle\frac{\mathrm{d}f\big(x(t)\big)}{\mathrm{dt}}\right\rangle \;=\; \frac{\mathrm{d}}{\mathrm{d}t}\left\langle f\big(x(t)\big)\right\rangle = \\ &=\; \left\langle a\big(x(t),t\big)\frac{\partial f\big(x(t)\big)}{\partial x} \;+\; \frac{1}{2}\,b\big(x(t),t\big)\frac{\partial^2 f\big(x(t)\big)}{\partial x^2}\right\rangle\;.\end{aligned}$$

The stochastic variable $\mathcal{X}(t)$ has the conditional probability density $p\,(x,t|\,x_0,t_0)$ and hence we can compute the expectation value by integration – again we simplify notation $f(x)\equiv f\big(x(t)\big)$ and $p\,(x,t)\equiv p\,(x,t|\,x_0,t_0)$:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left\langle f(x)\right\rangle \;&=\; \int \mathrm{d}x\, f(x)\,\frac{\partial}{\partial t}p\,(x,t) \;= \\ &=\; \int \mathrm{d}x\left(a(x,t)\,\frac{\partial f(x)}{\partial x} \;+\; \frac{1}{2}\,b(x,t)^2\,\frac{\partial^2 f(x)}{\partial x^2}\right)p\,(x,t)\end{aligned}$$

The further derivation follows the procedure that is used in the of the differential Chapman-Kolmogorov equation [157, 48-51] – in particular integration by parts and neglect of surface terms – and we obtain

$$\int \mathrm{d}x\, f(x)\,\frac{\partial}{\partial t}p\,(x,t) \;=\; \int \mathrm{d}x\, f(x)\Big(-\frac{\partial}{\partial x}\Big(A(x,t)\,p\,(x,t)\Big)+\frac{1}{2}\,\frac{\partial^2}{\partial x^2}\Big(B(x,t)^2\,p\,(x,t)\Big)\Big)\,.$$

Since the choice of a function $f(x)$ has been arbitrary we can drop it now and finally obtain a Fokker-Planck equation

$$\begin{aligned}\frac{\partial p\,(x,t|\,x_0,t_0)}{\partial t} &= -\frac{\partial}{\partial x}\Big(A(x,t)\,p\,(x,t|\,x_0,t_0)\Big) + \\ &\quad + \frac{1}{2}\frac{\partial^2}{\partial x^2}\Big(B(x,t)^2\,p\,(x,t|\,x_0,t_0)\Big)\ .\end{aligned} \tag{3.176}$$

The probability density $p\,(x,t)$ thus obeys an equation that is completely equivalent to the equation for a diffusion process characterized by a drift coefficient $a(x,t) \equiv A(x,t)$ and a diffusion coefficient $b(x,t) \equiv B(x,t)$ as derived from the Chapman-Kolmogorov equation. Hence, Itō's stochastic differential equation provides indeed a local approximation to a drift and diffusion process in probability space. The extension to the multidimensional case based on Itō's formula is straightforward, and we obtain for the conditional probability density $p\,(\mathbf{x},t|\mathbf{x}_0,t_0) \equiv p$ the following Fokker-Planck equation:

$$\frac{\partial p}{\partial t} = -\sum_i \frac{\partial}{\partial x_i}\big(A_i(\mathbf{x},t)\,p\big) + \frac{1}{2}\sum_{i,j}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}\Big(\big(\mathrm{B}(\mathbf{x},t)\cdot\mathrm{B}'(\mathbf{x},t)\big)_{i,j}\ p\Big)\ . \tag{3.177}$$

Here, we derive one additional property, which is relevant in practice. The stochastic differential equation,

$$\mathrm{d}\mathbf{x} = \mathbf{A}(\mathbf{x},t)\ \mathrm{dt} + \mathrm{B}(\mathbf{x},t)\,\mathrm{d}\mathbf{w}(t)\ , \tag{3.174'}$$

is mapped into a Fokker-Planck equation that depends only on the matrix product $\mathrm{B}\cdot\mathrm{B}'$ and accordingly, the same Fokker-Planck equation arises from all matrices B that give rise to the same product $\mathrm{B}\cdot\mathrm{B}'$. Thus, the Fokker-Planck equation is invariant to a replacement $\mathrm{B} \Rightarrow \mathrm{B}\cdot\mathrm{S}$ when S is an orthogonal matrix: $\mathrm{S}\cdot\mathrm{S}' = \mathbb{I}$. If S fulfils the orthogonality relation it may depend on $\mathbf{x}(t)$, but for the stochastic handling it has to be *nonanticipating*.

Eventually we proof the redundancy directly from the SDE and define a transformed Wiener process

$$\mathrm{d}\mathbf{v}(t) = \mathrm{S}(t)\,\mathrm{d}\mathbf{w}(t)\ .$$

The random vector $\mathbf{v}(t)$ is a normalized linear combination of Gaussian variables $\mathrm{d}w_i(t)$ and $\mathrm{S}(t)$ in nonanticipating, and accordingly, $\mathrm{d}\mathbf{v}(t)$ is itself Gaussian with the same correlation matrix. Averages $\mathrm{d}w_i(t)$ to various powers and taken at different times factorize and the same is true for the $\mathrm{d}v_i(t)$. Accordingly, the infinitesimal elements $\mathrm{d}\mathbf{v}(t)$ are increments of a Wiener process: The orthogonal transformation mixes trajectories without, however, changing the stochastic nature of the process, and equation (3.174) can be rewritten and yields

$$\begin{aligned} \mathrm{d}\mathbf{x} &= \mathbf{A}(\mathbf{x},t)\,\mathrm{dt} + \mathrm{B}(\mathbf{x},t)\,\mathrm{S}'(t)\cdot\mathrm{S}(t)\,\mathrm{d}\mathbf{w}(t) = \\ &= \mathbf{A}(\mathbf{x},t)\,\mathrm{dt} + \mathrm{B}(\mathbf{x},t)\,\mathrm{S}'(t)\cdot\mathrm{d}\mathbf{v}(t) = \\ &= \mathbf{A}(\mathbf{x},t)\,\mathrm{dt} + \mathrm{B}(\mathbf{x},t)\,\mathrm{S}'(t)\cdot\mathrm{d}\mathbf{w}(t)\ , \end{aligned}$$

since $\mathbf{v}(t)$ is as good a Wiener process as $\mathbf{w}(t)$ is, and both SDEs give rise to the same Fokker-Planck equation. □

3.4.5 Examples of stochastic differential equations

In order to show how stochastic differential equations can be handled in practice we show how to calculate first the expectation value and the variance of stochastic differential equations and then consider two cases: (i) the Ornstein-Uhlenbeck process that has been discussed as an example of a process that can be handled easily with a Fokker-Planck equation in section 3.2.2.3, and (ii) the general linear stochastic differential equations.

The Ornstein-Uhlenbeck process. The general SDE for the Ornstein-Uhlenbeck process has been given in (3.81). Without loosing generality but simplifying the solution we shift the long-time expectation value to the origin, $\mu = 0$:

$$\mathrm{d}x = -k\,x\,\mathrm{dt} + \sigma\,\mathrm{d}w(t)\ . \tag{3.81'}$$

The solution of the deterministic equation is simply and exponential decay or relaxation to the long-time value $\lim_{t\to\infty} x(t) = 0$,

$$\mathrm{d}x = -k\,x\ \mathrm{dt} \quad \text{and} \quad x(t) = x(0)\,e^{-kt}\ ,$$

and we make a substitution that compensates for the exponential decay

$$x(t) = y(t)\,e^{-kt} \quad \text{and} \quad y(t) = x(t)\,e^{kt} \quad \text{with} \quad y(0) = x(0)\ .$$

Now we expand $\mathrm{d}y$ up to second order

$$\mathrm{d}y = \mathrm{d}x\,e^{kt} + x\,\mathrm{d}(e^{kt}) + (\mathrm{d}x)^2 + \mathrm{d}x\,\mathrm{d}(e^{kt}) + \big(\mathrm{d}(e^{kt})\big)^2 \quad \text{with} \quad \mathrm{d}\big(e^{kt}\big) = ke^{kt}\,\mathrm{dt}\,.$$

All second order terms vanish because the expansion contains no term with $\mathrm{d}w(t)^2$ and we find by integration,

$$\mathrm{d}y = \sigma\,e^{kt}\,\mathrm{d}w(t) \quad \text{and} \quad y(t) = y(0) + \sigma\int_0^t e^{k\tau}\,\mathrm{d}w(\tau)\ ,$$

and resubstitution yields the solution

$$x(t) = x(0)\,e^{-kt} + \sigma\int_0^t e^{-k(t-\tau)}\,\mathrm{d}w(\tau)\ . \tag{3.178}$$

The calculation of expectation value and variance is straightforward:

$$\langle x(t)\rangle \;=\; \left\langle x(0)\,e^{-kt} \,+\, \sigma\int_0^t e^{-k(t-\tau)}\,\mathrm{d}w(\tau)\right\rangle \;=\; \langle x(0)\rangle\; e^{-kt}\;, \quad (3.179a)$$

and with

$$\begin{aligned}\left\langle x(t)^2\right\rangle \;&=\; \left\langle \left(x(0)\,e^{-kt} \,+\, \sigma\int_0^t e^{-k(t-\tau)}\,\mathrm{d}w(\tau)\right)^2\right\rangle \;= \\ &=\; \left\langle x(0)^2\right\rangle\, e^{-2kt} \,+\, \frac{\sigma^2}{2k}\left(1-e^{-2kt}\right)\end{aligned}$$

we obtain

$$\mathrm{var}\big(x(t)\big) \;=\; \left(\mathrm{var}\big(x(0)\big) - \frac{\sigma^2}{2k}\right) e^{-2kt} \,+\, \frac{\sigma^2}{2k}\;, \quad (3.179b)$$

and with sharp initial conditions, $p\,(x,0)=\delta(x-x_0)$, we find

$$\mathrm{var}\big(x(t)\big) \;=\; \left(\frac{1}{2k}\left(1-e^{-2kt}\right)\right)\;. \quad (3.179c)$$

Finally we mention that the analysis of the Ornstein-Uhlenbeck process can be readily extended to many dimensions and time dependent parameters, $k(t)$ and $\sigma(t)$ [157].

The linear stochastic differential equation. As last example we consider again the linear SDE but allow time dependent parameters

$$\mathrm{d}x \;=\; \alpha(t)\,x\,\mathrm{dt} \,+\, \beta(t)\,x\,\mathrm{d}w(t) \;=\; x\left(\alpha(t)\;\mathrm{dt} \,+\, \beta(t)\,\mathrm{d}w(t)\right)\;.$$

Now we make the substitution $y=\ln x$, expand up to second order

$$\mathrm{d}y \;=\; \frac{\mathrm{d}x}{x} - \frac{\mathrm{d}x^2}{x^2} \;=\; \alpha(t)\;\mathrm{dt} \,+\beta(t)\,\mathrm{d}w(t) \,-\, \frac{1}{2}\beta(t)^2\,\mathrm{dt}$$

and find the solution by integration and resubstitution

$$x(t) \;=\; x(0)\,\exp\left(\int_0^t\left(\alpha(\tau)-\frac{1}{2}\beta(\tau)^2\right)\mathrm{d}\tau \,+\, \int_0^t \beta(\tau)\,\mathrm{d}w(\tau)\right)\;. \quad (3.180)$$

We make use of the relation $\langle e^z\rangle = \exp\big(\frac{1}{2}\langle z^2\rangle\big)$, which is fulfilled by all Gaussian variables,[54] and find for the n-th raw moment [157, p. 109]:

[54] In order to proof the conjecture one makes use of the fact that all cumulants κ_n with $n>2$ vanish (see section 2.3.3). The reader encouraged to complete the proof.

$$\begin{aligned}\langle x(t)^n\rangle &= \langle x(0)^n\rangle \left\langle \exp\Big(n\int_0^t \big(\alpha(\tau) - \frac{1}{2}\beta(\tau)^2\big)\,\mathrm{d}\tau + n\int_0^t \beta(\tau)\,\mathrm{d}w(\tau)\Big)\right\rangle \\ &= \langle x(0)^n\rangle \exp\Big(n\int_0^t \alpha(\tau)\,\mathrm{d}\tau + \frac{1}{2}\,n(n-1)\int_0^t \beta(\tau)^2\,\mathrm{d}\tau\Big)\ .\end{aligned} \tag{3.181}$$

All moments can be calculated from this expression and for the low moments we find:

$$\langle x(t)\rangle = \langle x(0)\rangle \exp\Big(\int_0^t \alpha(\tau)\,\mathrm{d}\tau\Big) \tag{3.182a}$$

$$\begin{aligned}\mathrm{var}\big(x(t)\big) &= \mathrm{var}\big(x(0)\big)\exp\Big(2\int_0^t \alpha(\tau)\,\mathrm{d}\tau\Big) + \\ &\quad + \big\langle x(0)^2\big\rangle \exp\Big(\int_0^t \beta(\tau)^2\,\mathrm{d}\tau\Big)\ .\end{aligned} \tag{3.182b}$$

Analytical solutions have been derived also for the inhomogeneous case, $a(x,t) = \alpha_0 + \alpha_1 x$ and $b(x,t) = \beta_0 + \beta_1 x$ and the raw moments are readily calculated [157, p. 109].

Langevin equation in chemical kinetics. Although the chemical Langevin equation will be discussed extensively in section 4.2.3 we mention here a few fundamental properties already here. The conventional Langevin equation models a process in the presence of some random external force, which is expressed by the noise term $b\big(x(t),t\big)\mathrm{d}w(t)$. In order to keep the analysis simple we consider here only the case of a single reaction channel, and postpone reaction networks to chapter 4. In chemical kinetics such external forces may exist but the chemists are primarily interested in the internal fluctuations of particle numbers that ultimately result from the Poissonian nature of reaction events. Single reaction events are assumed to occur independently and the time interval between to reaction events is thought to follow an exponential distribution, and this implies that the total number of events denoted by m obeys a Poissonian distribution. In particular, if $\mathcal{P}(a,\tau)$ is the integer random variable denoting the number of reaction events that took place in the interval $t \in [0,\tau[$ the probability density, and $P_m(a,t) = P\big(\mathcal{P}(a,\tau) = m\big)$ with $a(\tau)$ being a function for the probability – or *propensity* – of the chemical reaction to take place. Then we have:[55]

$$P_m(a\tau) = \frac{(a\tau)^m}{m!}\,e^{-a\tau} \quad \text{with} \quad \mathrm{E}\big(\mathcal{P}(a\tau)\big) = a\tau \text{ and } \mathrm{var}\big(\mathcal{P}(a\tau)\big) = a\tau\ .$$

By the central limit theorem, or by using moment generating functions (section 2.3.3), or by direct computation making use of Stirling's formula we find,

[55] A Poissonian distribution depends only on a single parameter: $(a,\tau) \equiv (a\tau)$.

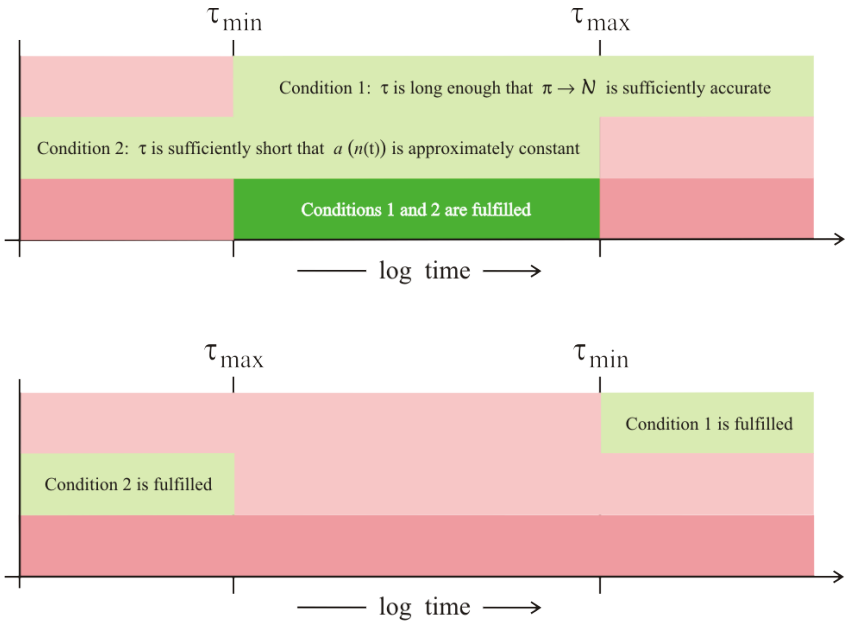


Fig. 3.30 Chemical Langevin equation. The chemical Langevin equation [171] is understood as an approximation to the master equation for modeling chemical reactions (section 4.2.1). The approximation is built upon two contradicting conditions concerning the time leap interval: (1) τ has to be sufficiently long in order to fulfil the relation $a\tau \gg 1$, and (2) τ has to be so short that the function $a\big(n(t)\big)$ does not change appreciably within the interval $[t, t+\tau]$. The upper part of the sketch shows a situation where the usage of the chemical Langevin equation is justified in the range $\tau_{\min} < \tau < \tau_{\max}$ whereas the Langevin equation is nowhere suitable under the conditions shown in the lower diagram.

$$\pi(a\tau) \doteq \frac{(a\tau)^k}{k!}\, e^{-a\tau} \approx \frac{1}{\sqrt{2\pi\, a\tau}}\, e^{-\frac{(k-a\tau)^2}{2a\tau}} \doteq \mathcal{N}(a\tau, a\tau) \quad \text{for} \quad a\tau \gg 1\,, \quad (2.40')$$

that the Poisson distribution can be approximated by a normal dustribution for large values of $a \cdot \tau$. This condition can be met either by large particle numbers or by long time intervals τ, or both (figure 3.30).

The number of molecules of species **A** at time t in the reaction volume V is modeled by the random variable $\mathcal{N}_{\mathbf{A}}(t) = n(t)$. Two relations from chemical kinetics are basic to stochastic modeling of the process [171]: (i) the chemical rate function or propensity function $a(n)$ for the reaction channel

$$a\big(n(t)\big)\ \mathrm{dt} \ \equiv \text{probability, given } n(t), \textbf{ one} \text{ reaction will occur within } [t, t+\ \mathrm{dt}[\quad (3.183a)$$

and (ii) the stoichiometric coefficient s, which determines the change in $\mathcal{N}_{\mathbf{A}}(t)$ by a single reaction event

$$\mathcal{N}_{\mathbf{A}}(t) = n \implies \mathcal{N}_{\mathbf{A}}(t + \mathrm{dt}) = n + s \text{ with } s \in \mathbb{Z} . \tag{3.183b}$$

It is worth noticing that in contrast to an ordinary Poisson process where we had $\Delta m = +1$ in a chemical reaction s might also be a negative integer, since molecules may be formed and may disappear, and need not be $s = \pm 1$ because, for example, molecules can be formed two or more at a time and they may also disappear two or more at a time (see section 4.1). The function $a(n)$ is the product of two factors:

$$a(n) = \gamma \cdot h(n) . \tag{3.183c}$$

The reaction rate parameter γ depends on external conditions like temperature, pressure, etc., does not depend on the number of particles or collision events, and in general is independent of time.[56] The particle number dependent factor $h(n)$ counts the number of events that can give rise to a reaction events, it may be simply $h(n) = n$ for spontaneous reactions involving one molecules or, for example, $n(n-1)/2$ if a collision of two molecules is required. Now we take recordings of the particle number in regular time intervals $\Delta t = \tau$. The number of reactions occurring in the time interval $[t, t+\tau]$ is another random variable denoted by $\mathcal{K}(n, \tau)$. For the time dependence of the number of molecules $\mathbf{A}$ we formulate the Markov process

$$\mathcal{N}_{\mathbf{A}}(t+\tau) = n(t+\tau) = n(t) + s\,\mathcal{K}_i(n, \tau) , \tag{3.183d}$$

which is equally hard to solve as the corresponding master equation. A straightforward approximation leading to a chemical Langevin equation, however, can be achieved provided two conditions are fulfilled:
Condition (1) requires that the time leap interval τ is sufficiently small that the chemical rate functions does not change appreciably

$$a\big(\mathcal{N}_{\mathbf{A}}(\theta)\big) \approx a\big(n(t)\big) , \ \forall\, \theta \in [t, t+\tau] \tag{3.184a}$$

Commonly the changes of particle numbers hardly exceeds $|s| = 2$ the condition is readily met by sufficiently large molecular populations as they are occurring in conventional chemical kinetics. Constancy of the rate functions at the same time guarantees independence for practical purposes of all reaction events within the interval $[t, t+\tau]$.
Condition (2) is required for the approximation of the Poissonian probability by a Gaussian distribution in the sense of equation (2.40'):

$$\mathrm{E}\Big(\mathcal{P}\big(a(n(t), \tau)\big)\Big) = a\big(n(t)\big)\,\tau \gg 1 . \tag{3.184b}$$

[56] Under certain rather rare circumstances modeling reactions with time dependent reaction rate parameters may be advantageous.

Apparently, the two conditions contradict each other and the existence of a range of validity for both conditions is not automatically given. In figure 3.30 we sketch two possible situation: The existence of a range of suitable value of τ fulfilling the two conditions where an approximation of the master equation by a Langevin-type equation is possible and a situation, where such a range does not exist.

Given the two conditions are fulfilled we can rewrite equation (3.183d) and find

$$\mathcal{N}_{\mathbf{A}}(t+\tau) \;=\; n(t) \;+\; s\mathcal{N}\Big(a\big(n(t)\big)\tau, a\big(n(t)\big)\tau\Big) \;.$$

Next we make use of the linear combination theorem of normal random variables, $\mathcal{N}(\mu,\sigma^2) = \mu + \sigma\mathcal{N}(0,1) = \mu + \sigma\varphi$ where $\varphi(x) = \exp(-x^2/2)/\sqrt{2\pi}$, and obtain

$$\mathcal{N}_{\mathbf{A}}(t+\tau) \;=\; n(t) \;+\; s\,a\big(n(t)\big)\,\tau \;+\; s\,\sqrt{a\big(n(t)\big)\,\tau}\,\mathcal{N}(0,1) \;.$$

The next step consists of an approximation in the same spirit of the condition (1) and (2): A time interval τ that fulfils the two conditions may be considered as *macroscopic infinitesimal* and therefore we treat τ as if it were a true infinitesimal dt, by replacing the discrete variable n by the continuous x, and by inserting $\mathrm{d}w = \varphi(t)\sqrt{\mathrm{dt}}$ we find:

$$\mathrm{d}x(t) \;=\; s\,a\big(x(t)\big)\;\mathrm{dt} \;+\; s\Big(a\big(x(t)\big)\Big)^{1/2}\,\mathrm{d}w(t) \;. \tag{3.185}$$

The chemical Langevin equation does not contain an external noise term but internal fluctuations resulting from the Poissonian nature of the chemical reactions events: $b(x(t),t) = s\,\sqrt{a\big(x(t)\big)}$, and the corresponding Fokker-Planck equation takes on the form

$$\frac{\partial P(x,t)}{\partial t} \;=\; -\,\frac{\partial}{\partial x}\Big(s\,a\big(x(t)\big)P(x,t)\Big) \;+\; \frac{\partial^2}{\partial x^2}\Big(s^2\,a\big(x(t)\big)P(x,t)\Big) \tag{3.186}$$

with the initial conditions (x_0, t_0).

We repeat what has been said before in this paragraph: The occurrence of chemical reaction events is modeled by a Poisson process but the consequences of the event itself follow the specific laws of chemical reactions. The conversion of a master equation into a Langevin equation is bound to the fulfillment of two contradictory conditions and under certain circumstances no acceptable Langevin approximation may be found. In this paragraph we have outlined only the basic features of modeling chemical reactions by means of Langevin equations. The formalism will be extended to reactions of several molecules and reaction networks in section 4.2.3 where we shall present applications as well.

Low moments of stochastic differential equations. In many cases it is sufficient to know the expectation value and the variance of the stochastic variable of a process as a function of time. These low moments[57] can be calculated without solving the stochastic equations explicitly. We consider the general SDE

$$\mathrm{d}x \;=\; a(x,t)\;\mathrm{dt} \;+\; b(x,t)\,\mathrm{d}w(t)$$

and compute the mean value by taking the average and recall that the second term on the r.h.s. vanishes because $\langle \mathrm{d}w(t)\rangle = 0$:

$$\mathrm{d}\,\langle x\rangle \;=\; \langle \mathrm{d}x\rangle \;=\; \langle a(x,t)\rangle\;\mathrm{dt} \quad \text{or} \quad \frac{\mathrm{d}\langle x\rangle}{\mathrm{dt}} \;=\; \langle a(x,t)\rangle\;. \tag{3.187}$$

Thus, the calculation of the expectation value boils down to solving an ODE. For a derivation of an expression for the second moment and the variance we have to calculate the differential of the square of the variable. By means of equation (3.172) we find:

$$\mathrm{d}(x^2) \;=\; \left(2x\,a(x,t) \;+\; b(x,t)^2\right)\;\mathrm{dt} \;+\; 2\,b(x,t)\,\mathrm{d}w(t)\;,$$

and forming the average yields

$$\left\langle \mathrm{d}(x^2)\right\rangle \;=\; \mathrm{d}\left\langle x^2\right\rangle \;=\left\langle 2x\,a(x,t) \;+\; b(x,t)^2\right\rangle\;\mathrm{dt}\;,$$

where we made use of the relation $\langle \mathrm{d}w(t)\rangle = 0$. Provided we knew the expectation values, a differential equation for the variance would be given by

$$\frac{\mathrm{d}\,\mathrm{var}(x)}{\mathrm{dt}} \;=\; \frac{\mathrm{d}\left\langle x^2\right\rangle}{\mathrm{dt}} \;-\; \frac{\mathrm{d}\langle x\rangle^2}{\mathrm{dt}} \;=\; \frac{\mathrm{d}\left\langle x^2\right\rangle}{\mathrm{dt}} \;-\; 2\,\langle x\rangle\;\frac{\mathrm{d}\langle x\rangle}{\mathrm{dt}}\;.$$

The continuation of the calculations requires knowledge of the functions $a(x,t)$ and $b(x,t)$.

As an example we consider the simple linear SDE with $a(x,t) = \alpha\,x$ and $b(x,t) = \beta\,x$,

$$\mathrm{d}x \;=\; \alpha\,x\,\mathrm{dt} \;+\; \beta\,x\,\mathrm{d}w(t) \;=\; x\left(\alpha\;\mathrm{dt} \;+\; \beta\,\mathrm{d}w(t)\right)\;,$$

and find for the expectation value

$$\langle x(t)\rangle \;=\; \langle x(0)\rangle\;e^{\alpha t} \;=\; x_0\,e^{\alpha t} \quad \text{for} \quad p\,(x,0) \;=\; \delta(x-x_0) \tag{3.188a}$$

and for the variance

[57] Low moments are expectation value and variance.

$$\begin{aligned} \mathrm{var}\big(x(t)\big) &= \big\langle x(t)^2 \big\rangle - \langle x(t) \rangle^2 = \\ &= \big\langle x(0)^2 \big\rangle \, e^{(2\alpha+\beta^2)t} - \langle x(0) \rangle^2 \, e^{2\alpha t} = \\ &= x_0^2 \left(e^{(2\alpha+\beta^2)t} - e^{2\alpha t} \right) \quad \text{for} \quad p\,(x,0) = \delta(x - x_0) \; . \end{aligned} \tag{3.188b}$$

The expressions are easily generalized to time dependent coefficients $\alpha(t)$ and $\beta(t)$ as we shall see in the paragraph on linear SDEs.

In this last part we have shown that analytical expressions derived from stochastic differential equations can be used successfully to compute the most important quantities of stochastic processes and in this sense are also equivalent to Fokker-Planck equations in practice.

Chapter 4
Applications in chemistry

There is nothing so practical as a good theory.
Kurt Lewin, 1952.

Abstract. In chemistry the master equation is the best suited and most commonly used tool to model stochasticity in reaction kinetics. We review the common elementary reactions in mass action kinetics and discuss Michaelis-Menten kinetics as an example of combining several elementary steps into an over-all reaction. Multistep reactions or reaction networks are considered and a formal mathematical theory that provides tools for the derivation of general properties of networks is presented. Then we digress into theory and empirical determination of rate parameters. The chemical master equation is introduced as the most popular tool for modeling stochasticity in chemical reactions, and the single reaction step approach is extended to reaction networks. Then, a selection of one-step reactions is presented for which the master equation can be solved exactly. The exact solutions are also used to illustrate the relation between the mathematical approach and the recorded data. A separate chapter is dealing with correlation functions, fluctuation spectroscopy, single molecule data and their stochastic modeling. Deterministic and stochastic parts of solutions can be separated by means of size expansions. Most reaction mechanisms are not accessible to the analytical approach and therefore we present a numerical approach that is exact within the concept of the chemical master equation is presented and applied to some selected examples of chemical reactions.

Conventional chemical reaction kinetics commonly does not require a stochastic approach because the numbers of particles are very large. There are exceptions when the particle numbers of certain species become very small during reactions – oscillations of species may serve as examples. Such cases will be mentioned and discussed in this and in the next chapter but even more important is the requirement of a stochastic approach for direct measurements of fluctuations, which became possible because of the progress in spectroscopy leading to spectacular increases in sensitivities. Single molecule techniques are another not completely unrelated and also rapidly developing field where a stochastic approach is indispensable. On the other hand, if one

wants to resolve reaction dynamics at the molecular level the situation is different, because conventional statistical mechanics is blurring the details of interest. Molecules are involved in large numbers of collisions, which considered individually in the vapor phase could be calculated at least in principle by means of advanced quantum mechanics, although we have to admit that a detailed computational approach to reactive collisions in solution where molecules are densely packed would be helpless.

Stochastic chemical kinetics is based on the assumption that knowledge on the transformation of molecules in chemical reactions is not accessible in full atomistic detail or if it would, the information would be overwhelming and obscuring the essential features. Thus, it is assumed that chemical reactions have a probabilistic element and can be modeled properly by means of stochastic processes. The random processes are caused by thermal noise as well as by random encounters of molecules in collisions. Fluctuations, therefore, play an important role and they are responsible for the limitations in the reproduction of experiments. This concept is not substantially different from the ideas underlying equilibrium statistical mechanics although statistics applied to thermodynamic equilibrium is on saver grounds than statistical mechanics applied to chemical reaction kinetics. On the other hand, the current theory of chemical reaction rates is around for more than fifty years and so far it has not yet been replaced by some better founded and applicable theory [270].

Particle numbers change necessarily in jumps requiring a discrete stochastic description, for example, by means of a master equation. Other descriptions are branching processes and other special cases of stochastic processes, which we will shall discuss in the next chapter 5, because they are more frequently addressed in biology. Commonly different approaches do not exclude each other as, for example, birth-and-death processes are frequently solved by application of precisely the same techniques as used for. Chemical master equations (section 4.2.1) and birth-and-death master equations, which were discussed in section 3.2.3.2, are closely related. Modeling chemical reactions by continuous variables as discussed in the paragraph on chemical Langevin equations (3.185) serves, in essence, two purposes: (i) It provides a natural transition from the stochastic to the deterministic treatment through increasing particle numbers $n(t)$ that reduces the contribution of the Wiener process, $s\sqrt{a\big(n(t)\big)}\mathrm{d}w$, until it becomes zero in the limit $n(t) \to \infty$, and (ii) it is indispensable for population size expansions, which provide the basis for a separation of the deterministic part of the solution from a diffusion term.

Conventional chemical reaction kinetics can be formally understood as the Liouville approach (section 3.2.2.1) to stochastic chemical kinetics and is dealing, in essence, with two classes of problems: (i) *forward problems*,[1]

[1] We remark that *forward* and *inverse* in the context of handling differential equations have nothing to do the direction of computational time in forward and backward Chapman-Kolmogorov equations.

which deal with the determination of time dependent concentrations as solutions of kinetic model equations, where the kinetic parameters are assumed to be known (for an introduction in to traditional chemical kinetics see [269], a modern textbook is [220]), and (ii) *inverse problems*, which aim at the determination of parameters from measured data, where the kinetic model is commonly assumed to be known [425]. The first problem boils down to deriving the solution curves or performing qualitatively analysis of a kinetic ODE, or a PDE in case the spatial distribution is nonhomogeneous. The inverse task is often addressed as *parameter identification problem*. Qualitative analysis of differential equations aims at the reconstruction of bifurcation patterns of dynamical systems and we encounter here an inverse problem too: The determination of the regions in parameter space from where parameter combinations give rise to a certain dynamic behavior [113].

The chapter starts with an introduction into conventional chemical kinetics (section 4.1) consisting of short reviews of elementary step kinetics (section 4.1.1), Michaelis-Menten kinetics, which is discussed as an example of a reaction mechanism merging two or more single elementary steps into over-all reactions (section 4.1.2), and a formal theory of reaction networks conceived for the qualitative analysis of multidimensional kinetic differential equations (section 4.1.3). Then, we shall answer questions concerning the origin and the empirical determination of the rate parameters (sections 4.1.4 and 4.1.5). Stochasticity in chemical reactions is introduced in terms of the chemical master equation, we shall discuss how reaction networks can be analyzed stochastically, and come back to chemical Langevin equations through handling multidimensional cases (section 4.2). Then, a collection of examples of exactly solvable chemical master equations is presented: (i) the equilibration of particle numbers or concentrations in the flow reactor, (ii) irreversible and reversible monomolecular reactions, and (iii) bimolecular reactions that can be still solved exactly but where the solutions become so complicated that practical work with them has to rely on numerical computation (section 4.3). A separate chapter is dealing with correlation functions, fluctuation spectroscopy, single molecule techniques and their implications as challenges for stochastic modeling (section 4.4). The next section deals with the transition from microscopic to macroscopic systems by means of power series expansions in an extensive physical parameter Ω, for example the size of the system or the total number of particles. Size expansion is particularly useful if the particle numbers are sufficiently large (section 4.5). Most reaction mechanisms involve many reactions steps and commonly exact analytical solutions are neither available for the conventional deterministic approach nor for stochastic methods. The last sections deal with a numerical approach to stochastic chemical kinetics in which probability distribution or low moments are obtained by sampling a sufficiently large number of numerically calculated individual trajectories (section 4.6).

4.1 A glance on chemical reaction kinetics

Chemical reactions will be modeled as Markov processes and analyzed in form of the corresponding master equations. In a few cases Langevin or Fokker-Planck equations will be applied too. Commonly, a chemical reaction mechanism is a network of several reaction steps. A reaction step will be called *elementary* if no further resolution is possible at the level of atoms or molecules.[2] Appropriate criteria for the classification of elementary steps are the molecularity of reactions[3] and the complexity of the reaction dynamics. The molecularity is discussed in the next section 4.1.1. With respect to reaction dynamics we shall consider reactions and reaction networks with (i) linear behavior, (ii) nonlinear behavior with simple dynamics in the sense of a monotonous approach to thermodynamic equilibrium or towards a unique stationary state, and (iii) complex behavior as exhibited by dynamical systems showing multiple stable stationary states, oscillations or deterministic chaos.

The stochastic approach to chemical reaction kinetics has some tradition, which began in the late fifties from two different initiatives: (i) approximation of the complex vibrational relaxation in small molecules [32, 335, 409] and its application to chemical reactions, and (ii) modeling of chemical reactions as stochastic processes [27, 28, 29]. The latter approach can be viewed in the sense of initially mentioned limited information on reaction details and has been taken up and developed further by several groups [85, 84, 225, 249, 273, 315, 318]. The major part of the early works has been summarized in an early review [316], which is recommended here for further reading. Anthony Bartholomay's studies are also highly relevant for biological models of evolution, because he modeled reproduction as a linear birth-and-death process. Exact solutions to master equations or Langevin and Fokker-Planck equations can be derived only for particularly simple special cases. Often approximations are used or the analysis is been restricted to expectation values and variances of the variables. Later on computer assisted approximation techniques and numerical simulation methods were developed, which allow for handling stochastic phenomena in chemical kinetics in a more general manner [157, 173, 439].

---

[2] In modern spectroscopy further resolution into different molecular or atomic states can be achieved and then the different states have be treated as individual entities in reaction kinetics (see equation (4.46)). A simple example of such a higher resolution is applied in modeling monomolecular reactions (section 4.1.4).

[3] The *molecularity* of a reaction is the number of molecules that are involved in the reaction, for example two in a reactive collision between molecules or one in a conformational change. An elementary step is a reaction at the molecular level that cannot be resolved further in mass action kinetics (section 4.1.1). We shall distinguish elementary steps and elementary processes: the latter are more general and need not be referring to the level of molecules.

4.1.1 Elementary steps of chemical reactions

Chemical reactions at the level of mass action kinetics are defined by mechanisms, which can be decomposed into *elementary steps*. Elementary steps describe the transformation of reactant molecules into products and are written as stoichiometric equations.[4] Common elementary steps involving zero, one or two reactant molecules are:

$$\star \longrightarrow \mathbf{A} \tag{4.1a}$$

$$\mathbf{A} \longrightarrow \oslash \tag{4.1b}$$

$$\mathbf{A} \longrightarrow \mathbf{B} \tag{4.1c}$$

$$\mathbf{A} \longrightarrow 2\,\mathbf{B} \tag{4.1d}$$

$$\mathbf{A} \longrightarrow \mathbf{B} + \mathbf{C} \tag{4.1e}$$

$$\mathbf{A} + \mathbf{B} \longrightarrow \mathbf{C} \tag{4.1f}$$

$$\mathbf{A} + \mathbf{B} \longrightarrow 2\,\mathbf{A} \tag{4.1g}$$

$$\mathbf{A} + \mathbf{B} \longrightarrow \mathbf{C} + \mathbf{B} \tag{4.1h}$$

$$\mathbf{A} + \mathbf{B} \longrightarrow \mathbf{C} + \mathbf{D} \tag{4.1i}$$

$$2\,\mathbf{A} \longrightarrow \mathbf{B} \tag{4.1j}$$

$$2\,\mathbf{A} \longrightarrow 2\,\mathbf{B} \tag{4.1k}$$

$$2\,\mathbf{A} \longrightarrow \mathbf{B} + \mathbf{C} \tag{4.1l}$$

The molecularity of a reaction is defined by the number of – different or identical – molecules on the reactant side of the stoichiometric equation and we distinguish *zero-*, *mono-*, *bi-*, or *termolecular*, reactions and so on.

The list shown above contains one zero-molecular reaction, (4.1a), four monomolecular reactions, (4.1b) - (4.1e), and seven bimolecular reactions, (4.1f) - (4.1l). Nonreactive events, which occur in open systems, for example in flow reactors, are the creation of molecules through inflow (4.1a)[5] or the annihilation of molecules through outflow (4.1b). They were included in the list because we shall need them to conceive examples of open systems. Elementary steps with molecularities of three and higher are not included in the list, because simultaneous encounters of three and more molecules are extremely improbable and apart from exceptions elementary steps involving

[4] Stoichiometry deals with the relative quantities of reactants and products in chemical reactions. Reaction stoichiometry, in particular, determines the molar ratios of the reactants, which are converted into products, and the products that are formed. For example, in the reaction $2\mathbf{H}_2 + \mathbf{O}_2 \rightarrow 2\mathbf{H}_2\mathbf{O}$ the stoichiometric ratios of $\mathbf{H}_2 : \mathbf{O}_2 : \mathbf{H}_2\mathbf{O}$ are $2 : 1 : 2$ (see also the stoichiometric matrix defined in section 4.1.3).

[5] The simplest kind of inflow as indicated here by the *star* is an arrival process with independent events and follows a Poissonian probability law.

three or more molecules are not considered in conventional chemical kinetics therefore.[6]

All elementary steps in the list with two exceptions being the equations (4.1g) and (4.1h) are characterized by the fact that molecular species represent either reactants or products. In other words, no molecules show up at both sides of the arrow. These two exceptions are: (i) catalysis (4.1h) and (ii) autocatalysis (4.1g). Both processes have specific features that make them unique among other chemical reactions. Catalysis implies the presence of a catalyzer, **B** in reaction (4.1h), which is consumed and produced in same amounts during the reaction. The presence of more catalyst increases the rate of the reaction. Catalysis by protein enzymes is the basis of biochemical kinetics and it is commonly described by multistep reactions, the Michaelis-Menten mechanisms discussed in section 4.1.2 being the most popular example. The elementary step shown in equation (4.1g) is an example of an *autocatalytic* elementary process. The unique feature of autocatalysis is self-enhancement of chemical species, $\mathbf{X} \rightarrow 2\mathbf{X}$, leading to amplification of fluctuations and exponential growth of particle numbers in case of (4.1g). In practice, autocatalytic reactions involve almost always many elementary steps and obey complex reaction mechanisms (see, e.g., the review [387]). The formal kinetics of reproduction meets the conditions of autocatalysis, and in this case again a complex multistep process is subsumed in a one-step over-all reaction, here of type (4.1g). Because of its fundamental importance of in biology we shall discuss autocatalysis in section 5.1 within the chapter dealing with applications in biology.

In order to model and analyze basic features of autocatalysis or chemical self-enhancement, single step autocatalytic reactions rather than autocatalytic multistep reaction networks are used in case studies. Despite its termolecular nature, one particular termolecular autocatalytic process,

$$\mathbf{A} + 2\,\mathbf{X} \longrightarrow 3\,\mathbf{X}\,, \tag{4.1m}$$

became very popular [346], although it unlikely to occur in pure form in real systems. The elementary step (4.1m) is the essential step in the so-called Brusselator model, it can be straightforwardly addresses by rigorous mathematical analysis, and it gives rise to complex dynamical phenomena in space and time, which are otherwise rarely observed in chemical reaction systems. Among other features such special phenomena are: (i) multiple stationary states, (ii) chemical hysteresis, (iii) oscillations in concentrations, (iv) deterministic chaos, and (v) spontaneous formation of spatial structures. The last example is known as Turing instability [432] and is frequently used as a model for pattern formation or morphogenesis in biology [321]. An excellent review on *nonlinear phenomena* in chemistry is found in the literature [387].

[6] Such exceptions are reactions involving surfaces as third partner, which are important in gas phase kinetics and, for example, biochemical reactions involving macromolecules.

Stoichiometry and chemical equilibria. Although chemists were intuitively familiar with mass action throughout the nineteenth century, the precise formulation of a *law of mass action* (MA) is due to two Norwegians, the mathematician and chemist Cato Maximilian Guldberg and the chemist Peter Waage [454]. For reaction (4.1f), for example, mass action rate law yields

$$\frac{\mathrm{d}[\mathbf{A}]}{\mathrm{d}t} \;=\; \frac{\mathrm{d}[\mathbf{B}]}{\mathrm{d}t} \;=\; -\,\frac{\mathrm{d}[\mathbf{C}]}{\mathrm{d}t} \;=\; k\,[\mathbf{A}]\cdot[\mathbf{B}] \;=\; v^{(\mathrm{MA})}([\mathbf{A}],[\mathbf{B}])\;.$$

The rate of the reaction $v^{(\mathrm{MA})}$ is proportional to a rate parameter k and to the amounts $[\mathbf{A}]$ and $[\mathbf{B}]$ at which the two reactants **A** and **B** are present in the reaction system.[7] Depending on the nature of the reaction system the chemists prefer to use different units: (i) particle numbers N_{A} and N_{B} counting the numbers of molecules and being dimensionless, (ii) molar concentrations $c_{\mathsf{A}} = N_{\mathsf{A}}/(N_{\mathrm{L}}\,V)\,[c^0]$ and $c_{\mathsf{B}} = N_{\mathsf{B}}/(N_{\mathrm{L}}\,V)\,[c^0]$ (section 1.1) where $c^0 = 1\,[\mathrm{mol{\cdot}l^{-1}}]$ is used for the unit concentration, and (iii) dimensionless molar ratios $x_{\mathsf{A}} = N_{\mathsf{A}}/(N_{\mathsf{A}} + N_{\mathsf{B}})$ and $x_{\mathsf{B}} = N_{\mathsf{B}}/(N_{\mathsf{A}} + N_{\mathsf{B}})$ with $x_{\mathsf{A}} + x_{\mathsf{B}} = 1$, when only **A** and **B** are present in the reaction system.

Precisely, the law of mass action states that the rate of any given chemical reaction is proportional to the product of the concentrations or activities of the reactants.[8] In particular, the numbers of identical molecules that are consumed in a reaction step – called the stoichiometric coefficients ν_A and ν_B – appear as exponents of concentrations, $v^{(\mathrm{MA})}$ is the reaction rate, and k is a rate parameter:

$$\begin{aligned} &\nu_A\,\mathbf{A} + \nu_B\,\mathbf{B} \xrightarrow{\;k\;} \text{products} \quad\Longrightarrow \\ \Longrightarrow\quad &\text{reaction rate} = v^{(\mathrm{MA})} = -\frac{1}{\nu_A}\frac{\mathrm{d}a}{\mathrm{d}t} = -\frac{1}{\nu_B}\frac{\mathrm{d}b}{\mathrm{d}t} = k\,a^{\,\nu_A}\,b^{\,\nu_B}\;. \end{aligned} \tag{4.2}$$

In a reversible reaction,[9] which represents an acceptable chemical reaction in both directions and which can be understood as a special combination of two elementary steps compensating each other, the reverse reaction is accounted

---

[7] We shall use the notation with square brackets, $[\mathbf{A}]$ and $[\mathbf{B}]$, when we want to leave it open, which units are to be used.

[8] Several idealized regularities hold only in the limit of vanishing concentrations, $\lim c \to 0$. The idealized laws are retained through replacing concentrations by activities, $a_{\mathsf{A}} = [\mathbf{A}] = f_{\mathsf{A}} c_{\mathsf{A}}$. Unless stated otherwise we shall approximate activities by concentrations here and for the sake of simplicity use lower case letters to indicate the species: $f_{\mathsf{A}} \approx 1$ and $[\mathbf{A}] = a\,[\mathrm{mol{\cdot}l^{-1}}]$ and $f_{\mathsf{B}} \approx 1$ and $[\mathbf{B}] = b\,[\mathrm{mol{\cdot}l^{-1}}]$.

[9] The notions *reversible* and *irreversible* for chemical reactions are used differently from the notions in thermodynamics: In chemical kinetics a reaction is *irreversible* if the occurrence of the reaction in opposite direction is not observable on realistic time scales and hence can be neglected. Strict chemical irreversibility causes an instability in thermodynamics. All chemical reactions that proceed with nonzero velocity are *irreversible* in the sense of thermodynamics as reversibility requires infinitely slow progress of processes and chemical reactions are no exception.

for by a negative sign and we obtain for the case of disjunct reactants and products:

$$\begin{aligned} & \nu_A \mathbf{A} + \nu_B \mathbf{B} \underset{l}{\overset{k}{\rightleftharpoons}} \nu_C \mathbf{C} + \nu_D \mathbf{D} \quad \Longrightarrow \\ \Longrightarrow \quad & v^{(\mathrm{MA})} = -\frac{1}{\nu_A}\frac{\mathrm{d}a}{\mathrm{d}t} = -\frac{1}{\nu_B}\frac{\mathrm{d}b}{\mathrm{d}t} = \frac{1}{\nu_C}\frac{\mathrm{d}c}{\mathrm{d}t} = \frac{1}{\nu_D}\frac{\mathrm{d}d}{\mathrm{d}t} = \\ & = v_{\rightarrow} - v_{\leftarrow} = k\, a^{\nu_A}\, b^{\nu_B} - l\, c^{\nu_C}\, d^{\nu_D}\,. \end{aligned} \tag{4.3}$$

The condition of zero net reaction rate yields an expression for the equilibrium parameter, commonly called the *equilibrium constant*, as already pointed out in the formulation of mass action at equilibrium by Guldberg and Waage:

$$K = \frac{k}{l} = \frac{\bar{c}^{\nu_C}\, \bar{d}^{\nu_D}}{\bar{a}^{\nu_A}\, \bar{b}^{\nu_B}}\,, \tag{4.4}$$

where the *bar* denotes equilibrium concentrations. Later derivations of mass action are using the chemical potentials of reactants and products and were first introduced by Josiah Willard Gibbs around nineteen hundred [162, 163].

In order to generalize the expressions we introduce M chemical species in a single reaction step[10] and allow species to appear on both sides of the reaction arrows:

$$\nu_1 \mathbf{X}_1 + \nu_2 \mathbf{X}_2 + \ldots + \nu_M \mathbf{X}_M \underset{l}{\overset{k}{\rightleftharpoons}} \nu_1' \mathbf{X}_1 + \nu_2' \mathbf{X}_2 + \ldots + \nu_M' \mathbf{X}_M\,. \tag{4.5}$$

In the following we shall use column vector notation for the concentrations, $\mathbf{x} = (x_1, \ldots, x_M)^{\mathrm{t}}$ and for the stoichiometric coefficients, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_M)^{\mathrm{t}}$. Then the rate functions take on the form

$$\begin{aligned} v_{\rightarrow}(\mathbf{x}) &= k\, x_1^{\nu_1} \cdot x_2^{\nu_2} \cdot \ldots \cdot x_M^{\nu_M} = k\, \mathbf{x}^{\boldsymbol{\nu}} \quad \text{and} \\ v_{\leftarrow}(\mathbf{x}) &= l\, x_1^{\nu_1'} \cdot x_2^{\nu_2'} \cdot \ldots \cdot x_M^{\nu_M'} = l\, \mathbf{x}^{\boldsymbol{\nu}'} \end{aligned} \tag{4.6}$$

where we use *multi-index notation* $\mathbf{x}^{\boldsymbol{\nu}} = (x_1^{\nu_1} x_2^{\nu_2} \cdots x_M^{\nu_M})$, and where unprimed and primed coefficients coefficients, $\boldsymbol{\nu}$ and $\boldsymbol{\nu}'$, refer to the reactant and product side, respectively. At equilibrium we have we have $\bar{v} = \bar{v}_{\rightarrow} = \bar{v}_{\leftarrow}$, or $k\bar{\mathbf{x}}^{\boldsymbol{\nu}} = l\bar{\mathbf{x}}^{\boldsymbol{\nu}'}$. The stoichiometric coefficients are reformulated by accounting for the net production of a compound in a reversible reaction: $s_j = \nu_j' - \nu_j$, which yields the differential equation

[10] Later on we shall be dealing with multistep reaction networks of irreversible and reversible reactions and apply a notation that allows for straightforward identification of reaction steps by choosing k_j and l_j as reaction parameters for reaction $\mathbf{R}_j$. The stoichiometric coefficients of the reactants in the reaction $\mathbf{R}_j$ will be denoted by $\nu_{Aj}, \nu_{Bj}, \ldots$, for the reaction products we shall use $\nu'_{Aj}, \nu'_{Bj}, \ldots$ and the elements of the stoichiometric matrix are $\mathrm{S} = \{s_{ij} = \nu'_{ij} - \nu_{ij}\}$ (see section 4.1.3).

$$\frac{1}{s_j}\,\frac{\mathrm{d}x_j}{\mathrm{dt}} \;=\; v_{\rightarrow}(\mathbf{x}) \;-\; v_{\leftarrow}(\mathbf{x})\;, \quad \forall\; j = 1, \ldots, M\;.$$

This equation is equally valid for a reversible reactions as well as for both irreversible reactions related to it, which result by setting either $l = 0$ or $k = 0$, respectively.

For the analysis of near equilibrium kinetics it is useful to define new variables that vanish at equilibrium,

$$\boldsymbol{\chi} \;=\; (\chi_1, \ldots, \chi_M)^{\mathrm{t}} \;=\; \mathbf{x} - \bar{\mathbf{x}} \;=\; (x_1 - \bar{x}_1, \ldots, x_M - \bar{x}_M)^{\mathrm{t}}\;,$$

and one common variable $\xi = \chi_i/(\nu_i' - \nu_i) = \chi_i/s_i \;\forall\; i = 1, \ldots, M)$.[11] Thermodynamics of irreversible processes requires that every process sufficiently close to the equilibrium state approaches stationarity by a simple exponential relaxation process.[12] After linearization around equilibrium we obtain

$$\frac{\mathrm{d}\xi}{\mathrm{dt}} \;=\; -\,\frac{1}{\tau_R}\;\xi \quad \text{and} \quad \xi(t) \;=\; \xi(0)\;\exp\!\big(-\,t/\tau_R\big)\;, \tag{4.7}$$

where τ_R is the so-called relaxation time of the chemical reaction [356, 471]

$$\tau_R^{-1} \;=\; \sum_{i=1}^{M}(\nu_i' - \nu_i)\frac{\nu_i\,\bar{v}_{\rightarrow} - \nu_i'\,\bar{v}_{\leftarrow}}{\bar{x}_i} \;=\; \sum_{i=1}^{M}\frac{(\nu_i' - \nu_i)^2\,\bar{v}}{\bar{x}_i}. \tag{4.8}$$

For the elementary steps in equation (4.1) the relaxation times are simple expressions, for example $\mathbf{A} \rightleftharpoons \mathbf{B} \;\rightarrow\; \tau_R^{-1} = k + l$, $2\mathbf{A} \rightleftharpoons \mathbf{B} \;\rightarrow\; \tau_R^{-1} = 4k\bar{a} + l$, $\mathbf{A} + \mathbf{B} \rightleftharpoons \mathbf{C} \;\rightarrow\; \tau_R^{-1} = k(\bar{a} + \bar{b}) + l$ or $2\mathbf{A} + \mathbf{B} \rightleftharpoons 3\mathbf{A} \;\rightarrow\; \tau_R^{-1} = (k(\bar{a} + \bar{b}) + l\bar{a})\bar{a}$. Relaxation in the multi-dimensional case will be discussed in section 4.1.3.

Equation (4.7) is the results of the linearization of a generally nonlinear expression. The nonlinear kinetic equation of a single step reaction can be cast into a straightforward integral equation

$$\mathbf{x}(t) \;=\; \mathbf{x}(0) \;+\; \left(\int_0^t v\Big(\mathbf{x}(\tau)\Big)\mathrm{d}\tau\right)(\boldsymbol{\nu}\,' - \boldsymbol{\nu}) = \;\mathbf{x}(0) \;+\; \left(\int_0^t v\Big(\mathbf{x}(\tau)\Big)\mathrm{d}\tau\right)\mathbf{s}\;, \tag{4.9}$$

The extension of the deterministic model to networks of arbitrary numbers of reactions is presented in section 4.1.3.

---

[11] The International Union of Pure and Applied Chemistry (IUPAC) has recommended to use the term *rate of reaction* exclusively for the differential quotient $\mathrm{d}\xi/\,\mathrm{dt} = |1/(v_i' - v_i)|(\mathrm{d}[\mathbf{X}_i]/\,\mathrm{dt})$, where ξ is the *degree of advancement* or the *extent of reaction* with the initial conditions as reference state: $\xi = ([\mathbf{X}_i] - [\mathbf{X}_i]_0)/s_i$. Here we define the variable ξ differently with the thermodynamic equilibrium as reference state. The variable ξ is independent of stoichiometric coefficients in both definitions.

[12] Linear laws near an equilibrium point are generally valid and not restricted to chemistry. Hook's law named after the English natural philosopher Robert Hook may serve as an example from mechanics.

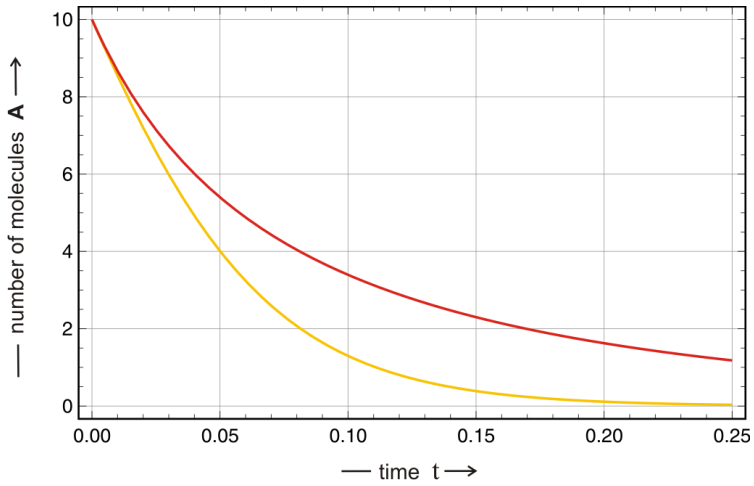


Fig. 4.1 The role of stoichiometry in kinetic equations. Reaction dynamics is determined by stoichiometry in an indirect way too. The two reactions shown in the figure correspond to the elementary step reactions (4.1i) and (4.1g), and follow formally the same rate law: $v = k\,[\mathbf{A}]\cdot[\mathbf{B}]$ and $v = k\,[\mathbf{A}]\cdot[\mathbf{X}]$. The two reactions differ in the stoichiometry on the product side, this leads to different conservation relations and further to different ODEs: $\mathrm{d}a/\,\mathrm{dt} = -k\,a\,(n_0 - a)$ and $\mathrm{d}a/\,\mathrm{dt} = -k\,a\,(\vartheta_0 + a)$, respectively, where n_0 and ϑ_0 are two different constants (For details see sections 4.1.3 and 4.3.3.3). Parameter choice: $k = 1.0$, $[\mathbf{A}]_{t=0} = 10$, $[\mathbf{B}]_{t=0} = [\mathbf{X}]_{t=0} = 15$. Color code: red, bimolecular conversion, and yellow, autocatalytic reaction.

It is important to point out that the product side exerts influence on the reaction dynamics also in an irreversible reaction if one or more reactants appear among the products. The best and simplest examples are autocatalytic reactions. For the purpose of illustration we compare the autocatalytic elementary reaction (4.1g) with the bimolecular conversion (4.1i). The kinetic differential equations and their solution are:

$$\frac{\mathrm{d}a}{\mathrm{dt}} = k\,a\,x = k\,a\,(n_0 - a) \quad \text{and} \quad a(t) = \frac{n_0\,a_0}{a_0 + (n_0 - a_0)\,\exp\,(n_0\,k\,t)}$$

$$\frac{\mathrm{d}a}{\mathrm{dt}} = k\,a\,b = k\,a\,(\vartheta_0 + a) \quad \text{and} \quad a(t) = \frac{\vartheta_0\,a_0}{(\vartheta_0 + a_0)\,\exp\,(n_0\,k\,t) - a_0}\,,$$

where a_0, x_0 and b_0 are the initial value of the variables a, x and b at time $t = 0$, $n_0 = a_0 + x_0$ and $\vartheta_0 = b_0 - a_0$, respectively.[13] In figure 4.1 we compare the two reactions with identical initial values of **A**, $a(0)$, and the

[13] Since neither **A** nor **B** appear on the product side it would make no difference to compare with (4.1f) or with $\mathbf{A} + \mathbf{B} \to 2\mathbf{C}$, which is the inversion of (4.1l).

same tangents, $\mathrm{d}a/\,\mathrm{dt}|_{t=0} = -k\,a_0\,x_0$ or $\mathrm{d}a/\,\mathrm{dt}|_{t=0} = -k\,a_0\,b_0$, respectively. We observe the build-up of a difference in rate that grows in time, which is due to self-enhancement of the autocatalyst: An increase in the concentration [**X**] leads to an increase in the reaction rate, a steady acceleration of the reaction, and faster consumption of **A**.

Another generalization in the notation of the differential reaction rate will turn out useful later on in handling chemical reactions as stochastic processes

$$\mathrm{d}\mathbf{x} \;=\; (v\;\mathrm{dt})\,\mathbf{s} \;=\; \big(k\,h(\mathbf{x})\;\mathrm{dt}\big)\;\mathbf{s}\;, \tag{4.10}$$

where k is the reaction parameter and the function $h(\mathbf{x})$ subsumes the concentration dependence of the differential change in the concentration vector $\mathbf{x}$. For the mathematical approach it is important that the reaction rate v is independent of dt and it is a scalar quantity expressing the fact that in a single reaction step there is one common reaction variable for all M molecular species. The function $h(\mathbf{x})$ contains the contribution of the concentrations of reactants and in mass action kinetics simply takes on the form $h(\mathbf{x}) = \mathbf{x}^{\boldsymbol{\nu}}$.

Strictly speaking, the resolution to the level of elementary steps implies the application of mass action kinetics, and this means that no further resolution is assumed to be achievable for molecules. As said, the advances in spectroscopy made it possible to distinguish between different states of molecules, in particular between ground states and various excited states in quantum molecular physics or between minimum free energy structures and suboptimal conformations in biopolymers, and then the ultimate resolution has to be pushed further down to individual states in order to be able to describe chemical reactions adequately.

Elementary step resolution and mass action kinetics often lead to complex reaction networks with a great number of variables, which are hard to analyze and which yield results that are difficult to interpret. It is sometimes useful to reduce the number of variables and to introduce some simpler *higher level kinetics*. The difference between mass action and higher level kinetics and is illustrated by means of an old and well studied example, the Michaelis-Menten reaction kinetics of enzyme catalyzed reactions in biochemistry.

4.1.2 Michaelis-Menten kinetics

Chemical kinetics became relevant for biology already at the end of the nineteenth century when biochemical processes gained a quantitative perspective. Biochemical kinetics became a discipline in its own right, and has been revived recently in the form of systems biology, which is raising the ambitious goal of modeling all processes in cells and whole organisms at the molecular level. In particular, enzyme catalyzed reaction were in the focus since the very beginning and indeed biochemical kinetics as we understand it today has been initiated by the path-breaking work of Leonor Michaelis and Maud Menten [329]: General enzyme catalysis is modeled by three elementary steps, which at first are assumed to be reversible:

$$\mathbf{S} + \mathbf{E} \rightleftharpoons \mathbf{S}\cdot\mathbf{E} \rightleftharpoons \mathbf{E}\cdot\mathbf{P} \rightleftharpoons \mathbf{E} + \mathbf{P}\,, \tag{4.11}$$

(i) binding of the substrate **S** to the enzyme **E**, (ii) conversion of substrate into product, both being bound to the enzyme, and (iii) the release of the product **P** through dissociation of the enzyme-product complex. Then, the full mechanism of simple enzyme catalyzed reaction consists of six elementary steps (figure 4.2):

$$\mathbf{S} + \mathbf{E} \xrightarrow{k_1} \mathbf{S}\cdot\mathbf{E}\,, \tag{4.12a}$$

$$\mathbf{S}\cdot\mathbf{E} \xrightarrow{l_1} \mathbf{S} + \mathbf{E}\,, \tag{4.12b}$$

$$\mathbf{S}\cdot\mathbf{E} \xrightarrow{k_2} \mathbf{E}\cdot\mathbf{P}\,, \tag{4.12c}$$

$$\mathbf{E}\cdot\mathbf{P} \xrightarrow{l_2} \mathbf{S}\cdot\mathbf{E}\,, \tag{4.12d}$$

$$\mathbf{E}\cdot\mathbf{P} \xrightarrow{k_3} \mathbf{P} + \mathbf{E}\,, \tag{4.12e}$$

$$\mathbf{P} + \mathbf{E} \xrightarrow{l_3} \mathbf{E}\cdot\mathbf{P}\,. \tag{4.12f}$$

For an efficient enzyme reaction it is essential that the steps (4.12d) and (4.12f) are negligibly slow. In particular, the latter reaction (4.12f) can lead to a substantial reduction of the production of the product **P** at high concentrations, a phenomenon known as *product inhibition* in biochemistry. It is necessary for high catalytic efficiency that reaction (4.12b) is slow too.

Here we present a brief analysis of the Michaelis-Menten mechanism by conventional chemical kinetics, and in sections 4.2.2 on stochastic reaction networks and 4.4 on single molecule techniques we shall come back to stochastic Michaelis-Menten kinetics with single enzyme molecules [295, 378].

Simple Michaelis-Menten kinetics. In the conventional Michaelis-Menten mechanism the reaction scheme (4.12) is simplified: Only the binding re-

action, step (i), is assumed to be reversible, whereas (ii) is thought to be an irreversible chemical reaction under the conditions of efficient enzyme catalysis. Step (iii) follows the irreversible reaction step (ii) and hence need not be considered explicitly:

$$\mathbf{S} + \mathbf{E} \underset{l_1}{\overset{k_1}{\rightleftharpoons}} \mathbf{S}\cdot\mathbf{E} \xrightarrow{k_2} \mathbf{E} + \mathbf{P}\,, \tag{4.13}$$

Simple Michaelis-Menten enzyme kinetics deals with four molecular species only, **S**, **E**, **S**·**E**, and **P** being substrate, enzyme, substrate-enzyme complex, and product, respectively. The enzyme-product complex, **E**·**P** is not considered explicitly, and the concentration of the product is interpreted as total concentration: $p \approx p_0 = [\mathbf{P}] + [\mathbf{E}\cdot\mathbf{P}]$. Again we denote concentrations by small letters, $[\mathbf{S}] = s$, $[\mathbf{E}] = e$, $[\mathbf{P}] = p$, and for the complex we use $[\mathbf{S}\cdot\mathbf{E}] = c_S = c$. Total concentrations will be denoted by: $e_0 = e + c$, $s_0 = s + c$ and p_0 as said above. In Michaelis-Menten kinetics ($v^{(\mathrm{MM})}$) the stoichiometric equations kinetic and the equation take on the form:

$$\mathbf{S} + \mathbf{E} \rightleftharpoons \mathbf{S}\cdot\mathbf{E} \to \mathbf{E} + \mathbf{P} \implies \text{reaction rate} = v^{(\mathrm{MM})} = \frac{\mathrm{d}[\mathbf{P}]}{\mathrm{dt}} = \frac{v_{\max}\cdot s}{K_M + s}\,.$$

The parameters $v_{\max}$ and K_M denote the maximal reaction rate and the Michaelis-Menten constant, respectively. The Michaelis-Menten constant is the free substrate concentration s at the half maximal reaction rate, $v_{\max}/2$. For more than half a century after the pioneering works of Michaelis and Menten, the Michaelis-Menten constant K_M has been the most important quantitative parameter of enzymes, and it has been used, for example, to determine the purity of enzyme preparations.

In order to derive the Michaelis-Menten equation we start from the mechanism given above. The differential equation for the enzyme substrate complex is of the form

$$\frac{\mathrm{d}c}{\mathrm{dt}} = k_1\, e\, s - (l_1 + k_2)\, c$$

and we obtain for the steady state:[14]

$$\frac{\mathrm{d}c}{\mathrm{dt}} = 0 \implies (l_1 + k_2)\,\hat{c} = k_1\,\hat{e}\,\hat{s}$$

Now we define the Michaelis-Menten constant and introduce $e_0 = e + c$ as the total enzyme concentration in order to eliminate the free enzyme concentration variable e:

[14] To indicate a true equilibrium state we would use the symbol $\bar{\cdot}$, e.g., $\bar{c}$. Since the assumption for the derivation of the Michaelis-Menten equation will be that the enzyme catalyzed reaction is sufficiently slow in order to keep the system in an approximate equilibrium state we are using $\hat{c}$, $\hat{s}$, etc., instead.

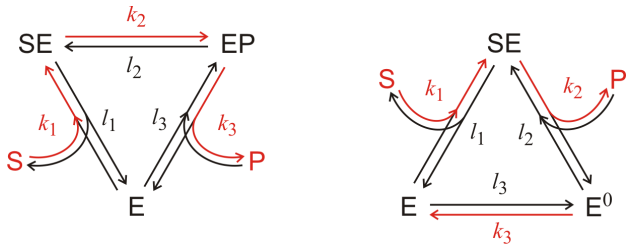


Fig. 4.2 The extended Michaelis-Menten mechanism. Two extended versions of the simplest Michaelis-Menten mechanism are consistent with empirical data of the majority of enzyme catalyzed reactions: (i) The mechanism on the l.h.s. (A) includes explicitly the enzyme-product complex **EP** and its dissociation into free enzyme **E** and product **P** (see equation (4.12) and, for example, [378]), and (ii) another extension of the simple Michaelis-Menten mechanism (4.13) includes an additional conformational state of the enzyme after release of the product from the complex, $\mathbf{E}^0$ (B). This mechanism is used, for example, in single molecule enzyme kinetics (see [262] and section 4.4.1). The highlighted path (red) illustrates the conversion of substrate **S** into product **P**.

$$\frac{l_1 + k_2}{k_1} = K_M = \frac{(e_0 - \hat{c})\,\hat{s}}{\hat{c}} \implies \hat{c} = \frac{e_0 \cdot \hat{s}}{K_M + \hat{s}} .$$

The rate of product formation is obtained through multiplication by the rate constant of the irreversible reaction

$$v^{(\mathrm{MM})} = \frac{\mathrm{d}p}{\mathrm{dt}} = k_2 \frac{e_0 \cdot s}{K_M + s} = \frac{v_{max} \cdot s}{K_M + s} \text{ with } v_{\max} = k_2\, e_0 , \quad (4.14)$$

and the result is the equation reported above. □

Often it is quite demanding to measure the free substrate concentration s and in the initial phase of the reaction or for $s_0 \gg e_0$, [**S**] can be approximated by the total substrate concentration $s \approx s_0$. An exact calculation is possible if the rate of reaction is zero and substrate binding is at equilibrium, $k_2 \ll l_1$:

$$\bar{s} = \frac{1}{2}\left((s_0 - e_0 - H) + (s_0 - e_0 + H)\sqrt{1 + \frac{4\,e_0 H}{(s_0 - e_0 + H)^2}}\right) , \quad (4.15)$$

with $H = l_1/k_1$ being the dissociation constant of the enzyme-substrate complex, $\mathbf{S} \cdot \mathbf{E}$. Equation (4.15) has a very simple solution under two conditions: (i) substrate **S** in large excess over enzyme **E**, $e_0 \ll s_0$ (and $e_0 \ll H$), and (ii) fast dissociation of the complex $l_1 \gg k_1$ or $\lim H \to \infty$

$$\bar{s} \approx s_0 - e_0 \approx s_0 .$$

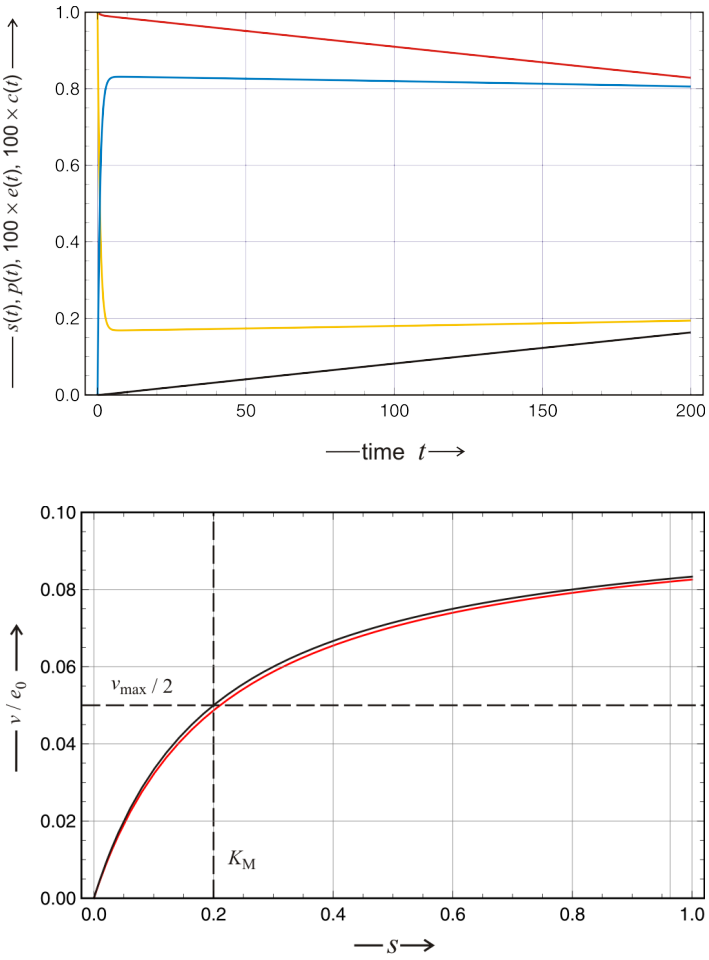


Fig. 4.3 The Michaelis-Menten mechanism for enzyme catalyzed reaction. The upper plot shows a numerical integration of the reaction scheme (4.13) leading to the ODE (4.17) with the parameters $k_1 = 1$, $l_1 = k_2 = 0.1$ and the initial concentrations $e_0 = 0.01$ and $s_0 = 1$. In the plot the concentrations $e(t)$ and $c(t)$ are multiplied by a factor 100. Color code: $s(t)$ red, $p(t)$ black, $e(t)$ yellow, and $c(t)$ blue. In the lower part of the figure the reaction rate $v = dp/dt = k_2(s_0 - s - p) = k_2(e_0 - e)$ is determined form a plot of v against s: v reaches a plateau value after an initial nonlinear increase, and this plateau value may be estimated from the maximum $v|_{dv/ds=0}$. The maximal rate is approximated by $v_{\max} = k_2(e_0 - e) \approx k_2 e_0$ because practically all enzyme **E** is converted into complex $\mathbf{S}\cdot\mathbf{E}$ at high substrate concentration, $s \gg e_0$. Choice of parameters: $k_1 = 1$, $l_1 = k_2 = 0.1$, $e_0 = 0.01$, and hence $K_M = 0.2$. The black curve $v(s)$ is compared with the plot of v against s_0 (red).

Without the equilibrium approximation Michaelis-Menten enzyme kinetics is described by two ODEs. The total concentrations of substrate and enzyme are according to stoichiometry

$$s(0) \;=\; s_0 \;=\; s + c + p\,,\; e(0) \;=\; e_0 \;=\; e + c\,,\;\; \text{and}\;\; c \;=\; e_0 - e\,, \tag{4.16}$$

where we have assumed that initially there was not product in the reaction mixture, $p\,(0) = p_0 = 0$:

$$\begin{aligned} \frac{\mathrm{d}p}{\mathrm{dt}} \;&=\; k_2\,(s_0 \,-\, s \,-\, p) \quad \text{and} \\ \frac{\mathrm{d}s}{\mathrm{dt}} \;&=\; -\,k_1\,s\cdot(e_0 \,-\, s_0 \,+\, s \,+\, p) \,+\, l_1\,(s_0 \,-\, s \,-\, p)\;. \end{aligned} \tag{4.17}$$

Results from computer integration of equation (4.17) are shown in figure 4.3. A fast binding reaction leading to a quasi-equilibrium is followed by relatively slow conversion of substrate into product that is characterized by an approximately constant concentration of the enzyme-substrate complex $\mathbf{S}\cdot\mathbf{E}$ that is tantamount to a constant rate of product synthesis. The Michaelis-Menten constant is obtained straightforwardly from the substrate concentration s at half-maximal reaction rate $v_{\max}/2$. It is also worth noticing how small the differences between s and s_0 are in this particular case.

The most important results of the Michaelis-Menten analysis of enzyme catalyzed reactions are: (i) A small value of the Michaelis-Menten constant K_M means that the enzyme reaches its maximal turnover already at small substrate concentrations, (ii) a large value of K_M implies the opposite – the maximal reaction rate is achieved only at high substrate concentrations, and (iii) the Michaelis-Menten constant K_M is proportional to the sum $l_1 + k_2$ and therefore large K_M does not necessarily imply a high catalytic rate parameter $k_2 = k_{\mathrm{cat}}$, it can also indicate weak binding of the substrate. Michaelis-Menten kinetics saw a recent revival when single-molecules studies of enzyme catalysis became possible (see section 4.4).

Extended Michaelis-Menten kinetics. Beginning in the nineteen sixties new spectroscopic and kinetic techniques were developed that allowed for resolution of reaction kinetics into individual reaction steps. The simple Michaelis-Menten mechanism (4.13) deals with only two states of the enzyme, free $\mathbf{E}$ and substrate bound $\mathbf{S}\cdot\mathbf{E}$, and gives rise to a single chemical relaxation mode (see section 4.4). Such simple kinetics is observed with few enzyme catalyzed reactions only whereas most enzymes exhibit more complex kinetics with two or more relaxation modes [295, 378] or even oscillations [101]. Based on this empirical evidence two extended versions of the original Michaelis-Menten mechanism are in use (figure 4.2, A and B). We shall apply here the natural extension by the explicit consideration of the enzyme-product complex shown in (4.12) and find for the five kinetic equations using $[\mathbf{S}\cdot\mathbf{E}] = c$ and $[\mathbf{E}\cdot\mathbf{P}] = d$ for the protein substrate complexes:

$$\frac{\mathrm{d}e}{\mathrm{d}t} = -(k_1' s + l_3' p)\, e + l_1\, c + k_3\, d\ , \tag{4.18a}$$

$$\frac{\mathrm{d}c}{\mathrm{d}t} = -(k_2 + l_1)\, c + k_1' s\, e + l_2\, d\ , \tag{4.18b}$$

$$\frac{\mathrm{d}d}{\mathrm{d}t} = -(k_3 + l_2)\, d + k_2\, c + l_3' p\ e\ , \tag{4.18c}$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -k_1'\, s\, e + l_1\, c\ , \quad \text{and} \tag{4.18d}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -l_3'\, p\ e + k_3\, d\ , \tag{4.18e}$$

where we choose primed symbols for the second order rate constants in order to facilitate the forthcoming change in notations: $k_1 = k_1' s$ and $l_3 = l_3' p$. The concentrations in the mechanism (4.18) converge to a thermodynamic equilibrium (see section 4.2.2)

$$\frac{\bar{p}}{\bar{s}} = \frac{[\overline{\mathbf{P}}]}{[\overline{\mathbf{S}}]} = \frac{k_1'\, k_2\, k_3}{l_1\, l_2\, l_3'} = K_1\, K_2\, K_3\ . \tag{4.19}$$

The individual equilibrium concentrations of the extended Michaelis-Menten mechanism are readily computed, and for the initial condition zero product, $p\,(0) = 0$, the conservation relations are:

$$s(0) = s_0 = s + c + d + p\,, \quad \text{and} \quad e(0) = e_0 = e + c + d\ . \tag{4.20}$$

With $\alpha = 1 + K$, $\beta = K_1(1 + K_2)$, $K_1 = k_1'/l_1$, $K_2 = k_2/l_2$, $K_3 = k_3/l_3'$, and $K = K_1 K_2 K_3$, we obtain:

$$\begin{aligned}
\bar{s} &= \frac{\beta(s_0 - e_0) - \alpha + \sqrt{4\alpha\beta s_0 + \Big(\beta(s_0 - e_0) - \alpha\Big)^2}}{2\,\alpha\beta}\ , \\
\bar{e} &= \frac{e_0}{1 + \beta\,\bar{s}}\ , \\
\bar{p} &= K_1 K_2 K_3\,\bar{s}\ , \\
\bar{c} &= K_1\,\bar{s}\,\bar{e}\ , \quad \text{and} \\
\bar{d} &= K_3^{-1}\,\bar{p}\,\bar{e}\ .
\end{aligned} \tag{4.21}$$

These expression for the equilibrium concentrations make them prohibitive for analytical work but they can be readily calculated numerically. The results, however, are mainly of academic interest since in the two cases of general importance the equilibrium conditions are never fulfilled in experiments: (i) If product formation is the goal, efficient synthesis under conditions far away from equilibrium are required, and (ii) in single molecule studies (see

section 4.4) the turnover of enzyme conformations occurs under conditions to which equilibrium thermodynamics cannot be applied. Numerical integration of equation (4.18) will be represented and discussed in section 4.6.4.

For many experimental investigations, in particular for single molecule experiments, the assumption of constant concentrations of substrate and product, $[\mathbf{S}] = s_0 = \text{const}$ and $[\mathbf{P}] = p_0 = \text{const}$, is realistic. Then the nonlinear ODE (4.18) is changed into a three dimensional linear ODE equation with $k_1 = k_1' \, s_0$ and $l_3 = l_3' \, p_0$:

$$\frac{\mathrm{d}}{\mathrm{dt}} \begin{pmatrix} e \\ c \\ d \end{pmatrix} = \begin{pmatrix} -(k_1 + l_3) & l_1 & k_3 \\ k_1 & -(k_2 + l_1) & l_2 \\ l_3 & k_2 & -(k_3 + l_2) \end{pmatrix} \cdot \begin{pmatrix} e \\ c \\ d \end{pmatrix} \ . \tag{4.22}$$

Now the analysis is straightforward [378], and the computation of the eigenvalues yields:[15]

$$\begin{aligned} \lambda_{1,2} &= -\frac{1}{2} \left((k_1 + k_2 + k_3 + l_1 + l_2 + l_3) \pm \sqrt{\Delta} \right) \quad \text{with} \\ \Delta &= (k_1 - k_2 - k_3 + l_1 - l_2 + l_3)^2 - 4\,(k_2 - l_3)(k_3 - l_1) \quad \text{and} \\ \lambda_3 &= 0 \ . \end{aligned} \tag{4.23}$$

The zero eigenvalue indicates a conservation relation that is given by the total enzyme concentration: $e_0 = e + c + d$. The commonly chosen experimental conditions apply no product, $[\mathbf{P}] = 0 \ \Rightarrow \ l_3 = 0$, or at least the initial condition $p\,(0) = 0$, and excess substrate, $[\mathbf{S}] = s \approx s_0 = [\mathbf{S}]_0$ where is the total concentration $[\mathbf{S}]_0$ is the sum of the concentrations of all complexes containing substrate or product: $s(0) = [\mathbf{S}]_0 + [\mathbf{P}]_0 = [\mathbf{S}] + [\mathbf{S} \cdot \mathbf{E}] + [\mathbf{E} \cdot \mathbf{P}] + [\mathbf{P}]$. The two nonzero eigenvalues are complex in the range [378]

$$h2 + \left(\sqrt{k2} - \sqrt{k3 - h1} \right)^2 < k_1 = k_1' [\mathbf{S}] < h2 + \left(\sqrt{k2} + \sqrt{k3 - h1} \right)^2 \ ,$$

and damped oscillations were indeed observed in single enzyme molecule experiments [101]. Damping of the oscillations is heavy because the ratio $\Im(\lambda)/\Re(\lambda)$ is small for this three state $(\mathbf{E}, \mathbf{S} \cdot \mathbf{E}, \mathbf{E} \cdot \mathbf{P})$ system.

[15] We remark that for a linear, n-dimensional ODE, $(\mathrm{d}/\,\mathrm{dt})\mathbf{x}^{\mathrm{t}} = \mathrm{A} \cdot \mathbf{x}^{\mathrm{t}}$, the matrix A is identical to the Jacobian matrix $\mathrm{J} = \{J_{ij} = \partial f_i / \partial x_j ;\, i, j = 1, \ldots, n\}$ for a general ODE, $(\mathrm{d}/\,\mathrm{dt})\mathbf{x}^{\mathrm{t}} = \mathbf{f}(\mathbf{x})^{\mathrm{t}}$, and its eigenvalues $\lambda_k ; k = 1, \ldots, n$ determine the (here global) stability of the system.

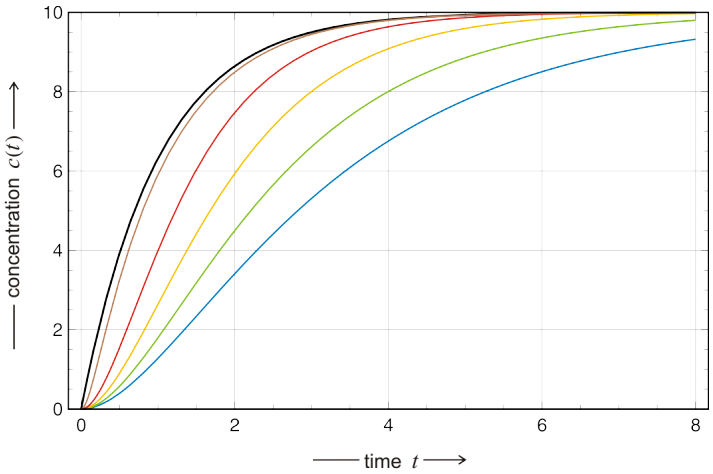


Fig. 4.4 The steady state approximation for multistep reactions. A test of the validity of the steady state approximation for the chain of irreversible first order reactions: **A** → **B** → **C** (4.25). The concentration of the reaction product **C** is plotted as a function of time. The steady state solution (black) becomes a better and better approximation of the exact curves the larger the value of k_2 is. Parameter choice: $a_0 = 10\,c^0$; $k_1 = 1\,[\text{t.u.}]^{-1}$; $k_2 = 0.4$ (blue), 0.6 (green), 1.0001 (yellow), 2 (red), and 10 (brown) $[\text{t.u.}]^{-1}$.

Generalized rate functions. It is often useful to define a common rate function for different kinetics, which allows for insertion of specific *association functions*:

$$v_j(\mathbf{x}) \;=\; k_j \prod_{i=1}^{M} \theta_i(x_i) \quad \text{with} \tag{4.24}$$

$$\theta_i(x_i) \;=\; x_i^{\nu_{i(j)}} \text{ for mass action kinetics , and}$$

$$\theta_i(x_i) \;=\; \frac{v_{\max}^{(ij)}\, x_i}{K_M^{(ij)} + x_i} \text{ for Michaelis-Menten kinetics ,}$$

where a combination of different kinetic function is possible.

There are many other forms of simplified kinetics in the sense that the number of independent variables is reduced at the expense of more complicated expressions. An example – closely related to the Michaelis-Menten approach – is the steady state approximation. We consider a two step reaction

$$\mathbf{A} \xrightarrow{k_1} \mathbf{B} \xrightarrow{k_2} \mathbf{C}\,, \tag{4.25a}$$

which is described by three kinetic differential equations. Only two are independent since we have the conservation relation $a + b + c = \text{const}$:

$$\frac{\mathrm{d}a}{\mathrm{d}t} \;=\; -k_1\, a\,, \quad \frac{\mathrm{d}b}{\mathrm{d}t} \;=\; k_1\, a \;-\; k_2\, b\,, \quad \text{and} \quad \frac{\mathrm{d}c}{\mathrm{d}t} \;=\; -k_2\, b \tag{4.25b}$$

The solution curves for the initial conditions $a(0) = a_0$ and $b(0) = c(0) = 0$ and $k_1 \neq k_2$ are

$$\begin{aligned} a(t) \;&=\; a_0\, e^{-k_1 t}\,, \; b(t) \;=\; a_0\, \frac{k_1}{k_2 - k_1} \left(e^{-k_1 t} - e^{-k_2 t}\right)\,, \quad \text{and} \\ c(t) \;&=\; a_0 \left(1 + \frac{k_1 e^{-k_2 t} - k_2 e^{-k_1 t}}{k_2 - k_1}\right)\,. \end{aligned} \tag{4.25c}$$

The steady state approximation is based on the assumption that the concentration of the intermediate does not change: $\mathrm{d}b/\,\mathrm{d}t = 0$, which leads to $\hat{b}(t) = k_1\, a(t)/\, k_2$, and hence

$$c(t) \;\approx\; a_0 \left(1 - e^{-k_1 t}\right)\,. \tag{4.25d}$$

Figure 4.4 illustrates the validity of the steady state approximation as a function of the ratio k_2/k_1: the larger this ratio is the better is the agreement between approximation and exact solution. It is also worth considering the opposite situation, $k_1 \gg k_2$ than the limit yields

$$c(t) \;\approx\; a_0 \left(1 - e^{-k_2 t}\right) \quad \text{and} \quad b(t) \;\approx\; a_0\, e^{-k_2 t}\,. \tag{4.25e}$$

What we see here in this simple example is a manifestation of the rule of the rate determining step. The overall kinetics of a chain of reactions is determined by the slowest step called the rate determining step: This is step 1 for $k_2 \gg k_1$ and step 2 for $k_1 \gg k_2$.

4.1.3 Reaction network theory

So far we have considered chemical reactions either as single step processes or we discussed techniques that approximated multistep mechanisms by an *overall* single step.[16] Almost all interesting chemical systems, however, consist of networks of reactions that are characterized by a variety of interacting molecular species, and this leads to dynamical systems with more than one, often many variables, for which analytical solutions are available very rarely only.[17] In the second half of the twentieth century, when chemists and physicists began to consider kinetic differential equations as dynamical systems and started to apply qualitative analysis, new questions in addition to forward and inverse problems became relevant. These new questions are concerned with general properties of reaction networks, for example, to prove (i) whether or not a network can sustain multiple steady states in the positive orthant of concentration space, (ii) whether or not undamped oscillations resulting from a stable limit cycle are possible or (iii) whether of not a specific reaction network can display deterministic chaos. Some of these questions can be answered by the deterministic theory of chemical reaction networks described here, which is also the fundament for the stochastic approach. A complementary but also general technique that can be applied for finding answers to these questions consists in the inversion of qualitative analysis [113, 297, 296]: Inverse bifurcation analysis aims at an exploration of the domains in parameter space that give rise to certain forms of complex dynamics.

A formal deterministic theory of chemical reaction networks has been developed already in the nineteen seventieth by Fritz Horn, Roy Jackson, and Martin Feinberg [125, 218] in order to complement conventional chemical kinetics by providing tools that allow for the derivation of general results for entire classes of reaction networks. The theoretical approach became really popular only recently when chemical reaction kinetics has been applied to systems biology and it was realized that stochastic modeling of extended chemical reaction networks is required for any deeper understanding of regulation and control of cellular dynamics and cellular metabolism [72, 184]. Before we consider modeling of stochastic chemical reaction networks (SCRNs) in section 4.2.2 we present a brief introduction to the Feinberg-Horn-Jackson-theory, which allows for straightforward answers to otherwise difficult to predict properties of chemical reaction networks, for example, the nonexistence of multiple steady states or the absence of oscillating concentrations. The theory is not aiming at deducing the properties of networks for given sets of rate parameters but it derives tools for studying features of families of networks irrespectively of the particular choice of parameters.

---

[16] Two trivial exceptions were the inflow and outflow of a compound $\mathbf{A}$ in the flow reactor and the reversible reaction $\mathbf{A} \rightleftharpoons \mathbf{B}$. In both cases, however, we were dealing with a single stochastic variable counting the numbers of molecules $\mathbf{A}$.

[17] A exception was the two step irreversible reaction $\mathbf{A} \rightarrow \mathbf{B} \rightarrow \mathbf{C}$ (4.25a).

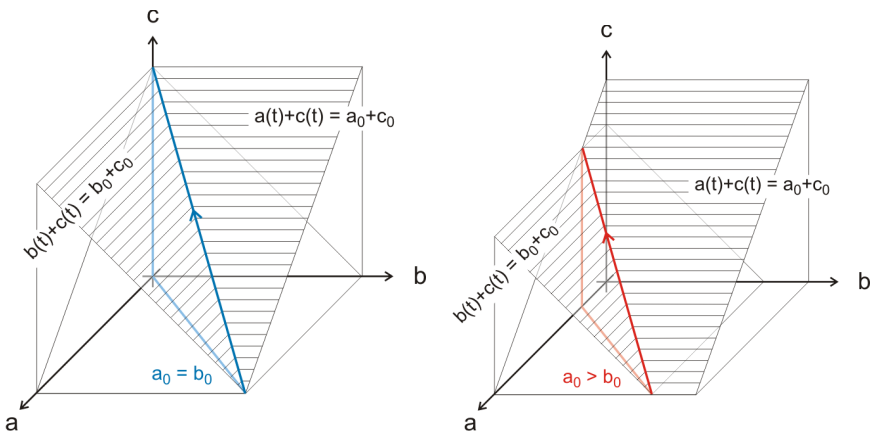


Fig. 4.5 Stoichiometric subspace and compatibility class. The figure on the r.h.s. sketches the stoichiometric subspace, $\mathbf{S} = \text{span}_j\{\mathbf{s}_j\}$, of the irreversible reaction $\mathbf{A}+\mathbf{B} \to \mathbf{C}$. The concentration space $\mathbf{X} = \{a, b, c\} \in \mathbb{R}^3$ is three dimensional, two independent conservation relations, $a(t) = a_0+c_0-c(t)$ and $b(t) = b_0+c_0-c(t)$, introduce linear dependencies and hence the stoichiometric subspace is one-dimensional. The stoichiometric compatibility class is formed by adding a constant vector $\mathbf{c} \in \mathbb{R}^M$, for example the initial conditions $\mathbf{x}_0 = (a_0, b_0, c_0)$ to the stoichiometric subspace: $\mathbf{x}_0+\mathbf{S}$. The two initial conditions applied here are: (i) $\mathbf{x}_0 = (a_0, b_0{=}a_0, 0)$ shown on the l.h.s., and (ii) $\mathbf{x}_0 = (a_0, b_0 < a_0, 0)$ on the r.h.s..

Formal stoichiometry. For the forthcoming discussions it is necessary to formalize the concept of stoichiometry in order to make it accessible to operations based on linear algebra. For this goal we assume a set of M chemical species $\boldsymbol{\Xi} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M\}$, which are interconverted by K chemical reactions, $\mathbf{R}_1, \mathbf{R}_2 \ldots, \mathbf{R}_K$. It is useful to define a row vector of species: $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M)$. Each individual chemical reaction $\mathbf{R}_j$

$$\sum_{i=1}^{M} \nu_{ij}\, \mathbf{X}_i \;\to\; \sum_{i=1}^{M} \nu'_{ij}\, \mathbf{X}_i \tag{4.26}$$

is characterized by two column vectors containing the stoichiometric coefficients $\boldsymbol{\nu}_j = \left(\nu_{1j}, \nu_{2j}, \ldots, \nu_{Mj}\right)^{\mathrm{t}}$ and $\boldsymbol{\nu}'_j = \left(\nu'_{1j}, \nu'_{2j}, \ldots, \nu'_{Mj}\right)^{\mathrm{t}}$ of reactants and products, respectively.Now we can write the stoichiometric equation of reaction $\mathbf{R}_j$ (4.26) in compact form

$$\mathbf{R}_j: \quad \mathbf{X} \cdot \boldsymbol{\nu}_j \;\to\; \mathbf{X} \cdot \boldsymbol{\nu}'_j \quad \text{and} \quad \mathbf{X} \cdot (\boldsymbol{\nu}'_j - \boldsymbol{\nu}_j) \;=\; \mathbf{X} \cdot \mathbf{s}_j \;. \tag{4.26'}$$

A linear combination of species as defined by the stoichiometry of a chemical reaction is characterized as *reaction complex* (section 4.1.3):[18] $\mathsf{C}_j = \mathbf{X} \cdot \boldsymbol{\nu}_j$ or $\mathsf{C}_{j'} = \mathbf{X} \cdot \boldsymbol{\nu}'_j$ being the reactant complex and the product complex of reaction $\mathbf{R}_j$, respectively. The stoichiometric coefficients of all N complexes appearing in a chemical reaction network together form the $M \times N$ *matrix of complexes*

$$\mathrm{C} \;=\; \big(\mathbf{X} \cdot \boldsymbol{\nu}_1 \;\; \mathbf{X} \cdot \boldsymbol{\nu}_2 \;\; \ldots \;\; \mathbf{X} \cdot \boldsymbol{\nu}_N\big) \;.$$

As indicated already in equation (4.26') we combine the stoichiometric vectors belonging to the reactants and the products of the same reaction whereby we count reactant coefficients as being negative in order to provide a measure of the change introduced by the reaction. The stoichiometry of the entire reaction network is properly encapsulated in the $M \times K$ *stoichiometric matrix*:

$$\mathrm{S} \;=\; (\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_K) \;=\; \{s_{ij}; i = 1, \ldots, M; j = 1, \ldots, K\} \tag{4.27}$$

The stoichiometric matrix allows for a compact written form of the kinetic differential equations and their solutions

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{dt}} \;=\; \mathrm{S} \cdot \mathbf{v} \;\; \text{and} \;\; \mathbf{x}(t) - \mathbf{x}_0 \;=\; \sum_{j=1}^{K} \left(\int_0^t v_j\big(\mathbf{x}(\tau)\big)\, \mathrm{d}\tau \right) \mathbf{s}_j \;, \tag{4.28}$$

where $\mathbf{v} = \Big(v_1\big(\mathbf{x}(t)\big), v_2\big(\mathbf{x}(t)\big), \ldots, v_K\big(\mathbf{x}(t)\big)\Big)^{\mathrm{t}}$ is the vector of reaction rates, here mass action rates $v^{(\mathrm{MA})}$ according to equation (4.3), the variables are concentrations described by a vector $\mathbf{x}(t) = \big(x_1(t), x_2(t), \ldots, x_M(t)\big) \in \mathbb{R}^M$ with $x_i(t) = [\mathbf{X}_i(t)]$ being the concentration of compound $\mathbf{X}_i$ at time t, and $\mathbf{x}_0 = \big(x_1(0), x_2(0), \ldots, x_M(0)\big)$ are the initial concentrations. Equation (4.28) is the straightforward extension of (4.9) to an arbitrary number of reactions.

A number of restrictions apply to chemical kinetics: (i) concentrations are positive real numbers, $x_j(t) \in \mathbb{R}_{>0}\, \forall\, j = 1, \ldots, M$,[19] (ii) the solutions have to fulfil the stoichiometric relations for all reactions $\mathbf{R}_j$ ($j = 1, \ldots, K$) and this is encapsulated in the restriction to *stoichiometric compatibility classes*. We define the *stoichiometric subspace* of a reaction system by

$$\mathbf{S} \;=\; \mathrm{span}\{\mathbf{s}_j \,|\, j = 1, \ldots, K\} \subset \mathbb{R}^M \;\; \text{and} \;\; R \doteq \dim(\mathbf{S}) \;. \tag{4.29}$$

---

[18] The notion of *reaction complex* needs affirmation, since it is different from an association complex like the enzyme-substrate complex in the Michaelis-Menten reaction: A reaction complex is a combination of molecules in the correct stoichiometric ratio as it appears at the reactant side or at the product side of a stoichiometric equation.

[19] In chemistry concentrations of molecular species are commonly required to be positive quantities, whereas extinction corresponding to concentration zero is often an important issue in biology. Then, *positive* has to be replaced only by *nonnegative*, $\mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$.

The stoichiometric compatibility class contains the stoichiometric subspace shifted by some constant vector, $\mathbf{c}+\mathbf{S}$, and we restrict the variables to positive values of the concentrations is of the form

$$\mathcal{D} \;=\; (\mathbf{c} \,+\, \mathrm{span}\{\mathbf{s}_j \,|\, j = 1, \ldots, K\}) \cap \mathbb{R}^M_{>0} \;=\; (\mathbf{c} \,+\, \mathbf{S}) \cap \mathbb{R}^M_{>0} \;. \qquad (4.30)$$

Figure 4.5 shows a simple example of a one-dimensional compatibility class embedded in a three-dimensional concentration space. Since the linear span is built from all reaction vectors $\mathbf{s}_j$, linear dependencies will occur in most cases. The number of independent vectors in $\mathrm{span}_j(\mathbf{s}_j)$, the dimension or the rank R of the stoichiometric subspace, is the number of independent concentration variables or the number of degrees of freedom in the kinetic reaction system. The rank R of the stoichiometric matrix represents the number of degrees of freedom of the kinetic system and is either determined analytically or computed by routine software. For small systems, like the examples presented here, it is useful and illustrative to reduce the degrees of freedom by means of easy to find conservation relations, but for larger system with several hundred variables and more, a stable numerical procedure is commonly to be preferred.

Chemical reaction networks. The notion of a chemical reaction network stands in the center of the reaction network theory. Each network consists of three commonly finite sets of objects

(i) a set of M molecular species, $\boldsymbol{\Xi} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M\}$, which interact through a finite number of chemical reactions,
(ii) a set of N complexes, $\mathcal{C} = \{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_N\}$, which are linear combinations of species, $\mathbf{C}_j = \sum_{i=1}^{M} \nu_{ij}\mathbf{X}_i$ with $\nu_{ij} \in \mathbb{N}_{>0}$, and
(iii) a set of K molecular reactions, $\mathcal{R} = \{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K\}$, with $\mathcal{R} \subset \mathcal{C} \times \mathcal{C}$ in the sense of individual elements being directed combinations of two complexes, $(\mathbf{C}_\mathrm{R}, \mathbf{C}_\mathrm{P}) \in \mathcal{R}$ is written as $\mathbf{C}_\mathrm{R} \to \mathbf{C}_\mathrm{P}$ where R and P stand for *reactants* or *products*, respectively.

Restrictions are imposed on the sets $\mathcal{S}$ and $\mathcal{C}$: Each element of $\mathcal{S}$ has to be found in at least one reaction complex or, in other words, there are no superfluous species. Condition (iii) is supplemented by two exclusions: No complex may react into itself, $\mathbf{C}_\mathrm{R} \neq \mathbf{C}_\mathrm{P}$, and isolated complexes are not allowed, in the sense that every element of $\mathcal{C}$ must be the reactant or the product complex of some reaction. It is worth reminding that a reversible reaction (see e.g. section 4.3.2.2) is represented by two reactions: $\mathbf{C}_\mathrm{R} \to \mathbf{C}_\mathrm{P}$ and $\mathbf{C}_\mathrm{P} \to \mathbf{C}_\mathrm{R}$.

The mentioned restriction can be cast in a somewhat different form that is presented here in order to clarify the definitions. Complexes and species are related through

(a) $\mathcal{C} \subset \mathbb{R}^\mathcal{S}$ where $\mathbb{R}^\mathcal{S}$ stands for a vector space spanned by unit vectors representing individual species. Commonly, the coefficients in the linear combinations of species called complexes are natural numbers, $s_{ij} \in \mathbb{N}> 0$,

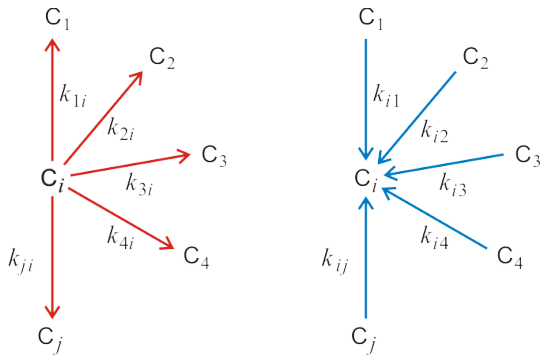


Fig. 4.6 Complex balancing. Balancing of complexes is achieved when the inflow into every complex $\mathbf{C}_i$ (blue) is precisely compensated by the outflow from it (red). Complex balancing is a relaxation of the constraint of *detailed balance* that requires that all individual reaction steps are at equilibrium.

(b) $\bigcup_{\text{supp, } \mathbf{C}_j \in \mathcal{C}} \mathbf{C}_j = \mathcal{S}$ the union of the species in all complexes is the species set and no species can exist in $\mathcal{S}$, which does not appear in at least one complex.[20]

Species $\mathbf{X}_i$ and reactions $\mathbf{R}_j$ are directly related by the stoichiometric matrix $\mathrm{S} = \{s_{ij}\}$. The columns of S refer to reactions and the rows to species. We shall make use of S also in section 4.6 for the implementation of a simulation tool for chemical master equations.

The fourth components of a reaction system is the kinetics of the reactions, $\mathcal{K}$. Mass action kinetics ($v^{(\mathrm{MA})}$) has been discussed in section 4.1.1 and Michaelis-Menten kinetics ($v^{(\mathrm{MM})}$) as an example of higher-level kinetics in section 4.1.2. In the majority of the examples discussed here mass action we be applied. We repeat the basic equation (4.2) for reaction $\mathbf{R}_j$:

$$s_{1j}\,\mathbf{X}_1 + s_{2j}\mathbf{X}_2 + \ldots \implies v_j = k_j\,[\mathbf{X}_1]^{s_{1j}} \cdot [\mathbf{X}_2]^{s_{2j}} \cdots = k_j\,x_1^{s_{1j}} \cdot x_2^{s_{2j}} \cdots . \quad (4.31)$$

In mass action kinetics $v^{(\mathrm{MA})}$ we need one reaction parameter, k_j, for every elementary step and hence the number of rate parameters is equal to K, the number of reactions.[21] Eventually, a reaction system consists of the four components $\{\mathcal{S}, \mathcal{C}, \mathcal{R}, \mathcal{K}\}$ and the evolution in time of the reaction system can

[20] The notion 'supp' stands for the support of a vector, which is the subset of unit vectors for which the vector has nonzero coefficients.

[21] In order to make the notation clearer for reversible reactions, we use two symbols and the same index for both reactions: k_j and l_j for the forward and the reverse reaction, respectively.

be encapsulated in an ODE or in a master equation in case of a stochastic description.

Stationary states. Two kinds of stationarity are important to distinguish: (i) equilibria with detailed balance and (ii) complex-balanced equilibria. Detailed balance follows from statistical thermodynamics and implies that the flow for every individual reaction step $\mathbf{R}_j$ vanishes at equilibrium [430]: $v_j = 0 = k_j \bar{x}_1^{s_{1j}} \cdot \bar{x}_2^{s_{2j}} \cdots \forall\, j = 1, \ldots, K$. It is realized the limit of chemical kinetics at thermodynamic equilibrium (3.100). The weaker condition of complex balancing [124, 217, 344] requires that for all complexes the net flow into a complex $\mathbf{C}_i$ is compensates by the net outflow (figure 4.6):[22]

$$\mathbf{C}_i : \ \frac{\mathrm{d}[\mathbf{C}_i]}{\mathrm{dt}} = 0 \quad \text{or} \quad \sum_{j,j\neq i}^{N} k_{ji}\, \widehat{\mathbf{c}}_i^{\,\boldsymbol{\nu}_j} = \widehat{\mathbf{c}}_i^{\,\boldsymbol{\nu}_i} \sum_{j,j\neq i}^{N} k_{ij} \ \ \forall \ \ \mathbf{C}_i \in \mathcal{C} \ . \qquad (4.32)$$

In order to facilitate the distinction equilibrium concentrations are characterized by a 'bar' and stationary concentrations obtained from complex balancing by a'hat'.

Reaction graphs. Some general properties of reaction networks can be predicted directly from the reaction graph (figure 4.7), which is a directed graph containing the *complexes*, $\mathbf{C}_k \in \mathcal{C}\,, (k = 1, \ldots, N)$, as nodes and three symbols indicating forward ($\rightarrow$), backward ($\leftarrow$) and reversible reaction ($\rightleftharpoons$) as edges. A reaction graph may have several components called *linkage classes*. Different linkage classes have no common node and no edge connecting them. Two properties are important for reaction graphs: (i) Every complex appears only once as a node of the graph and (ii) different linkage classes to not share complexes.

The network in figure 4.7 has two linkage classes since the two clusters don't share a single complex. The information on the number of complexes and the number of linkage classes is contained in the reaction graph. The same is true for the classification of a network as reversible, weak reversible or not reversible. A (strongly) reversible network contains exclusively reversible reactions in the strict thermodynamic sense. Weak reversibility relaxes the condition of (*strong*) *reversibility*: A network is weakly reversible when for every pair of complexes there exist a directed arc leading from one complex to the other. The network in figure 4.7 fulfils the condition of weak reversibility, it would be (strongly) reversible if it would be complemented by the arrows $\mathbf{C}_3 \rightarrow \mathbf{C}_5$ and $\mathbf{C}_5 \rightarrow \mathbf{C}_4$. For the determination of linkage classes only the existence or absence of arrows between complexes matters. Clearly, the direction of arrows is is required too for the classification of reversibility.

A reaction graph differs from a reaction mechanism in three aspects: The reaction complexes are not defined in terms of chemical compounds and there-

[22] For the definition and illustration of complex balancing it is convenient to apply a different notation of the rate parameter: For the reaction $\mathbf{C}_i \rightarrow \mathbf{C}_j$ we use k_{ji}.

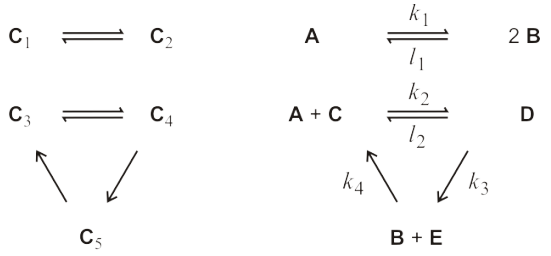


Fig. 4.7 The graph corresponding to the chemical reaction network (4.35a)**.** Each node of the graph (l.h.s.) corresponds to a reaction complex, three different symbols characterize the directed edges: $\rightarrow$, $\leftarrow$, and $\rightleftharpoons$ for *forward*, *backward*, and *reversible reaction*, respectively. This graph consists of $L = 2$ linkage classes. On the r.h.s. we show the Feinberg mechanism, which is an implementation of the reaction graph on the l.h.s. The mechanism differs from the graph by additional information: (i) the molecular realization of the reaction complexes and the rate parameters.

fore the reaction graph does not consider stoichiometry, it does not specify the algebraic relations of reaction rates in the form of mass action, Michaelis-Menten or other kinetic functions, and it does not contain weighting factors of edges in the sense of rate parameters. The reaction graph represents nothing more than the topology of a reaction network and general properties derived from the graph are valid for a large number of concrete cases irrespective of stoichiometries, kinetic functions, and rate constants.

Examples of reactions and networks. We illustrate chemical reaction network theory by means of a few examples.

The irreversible association reaction: $\mathbf{A} + \mathbf{B} \rightarrow \mathbf{C}$. [23] The first example is the irreversible association reaction (4.1f):

$$\mathbf{A} \; + \; \mathbf{B} \quad \xrightarrow{k} \quad \mathbf{C} \; . \tag{4.33a}$$

For the three sets of the chemical reaction network we have

$$\mathcal{S} \; = \; \{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \; , \tag{4.33c}$$

$$\mathcal{C} \; = \; \{\mathbf{C}_1 = \mathbf{A} + \mathbf{B}, \mathbf{C}_2 = \mathbf{C}\} \; , \quad \text{and} \tag{4.33d}$$

$$\mathcal{R} \; = \; \{\mathbf{R}_1 = \mathbf{C}_1 \rightarrow \mathbf{C}_2\} \; . \tag{4.33e}$$

The stoichiometric matrix S is of dimension 3×1:

[23] In narrative chemical kinetics distinctions are made for notions concerning the association-dissociation reaction $\mathbf{A} + \mathbf{B} \rightleftharpoons \mathbf{C}$ that are synonyms in formal kinetics: The word *addition* is used when **A** and **B** are of similar size and *binding* is preferred for molecules of very different size, for example, a substrate is *bound* to an enzyme.

$$\mathrm{S} \;=\; \begin{pmatrix} -1 \\ -1 \\ +1 \end{pmatrix} \;. \tag{4.33f}$$

In deterministic mass action kinetics, $v^{(\mathrm{MA})}$, the variables are the concentrations of the molecular species, $[\mathbf{A}] = a(t)$, $[\mathbf{B}] = b(t)$, and $[\mathbf{C}] = c(t)$. In order to solve the kinetic differential equation we require a rate parameters k and three initial conditions $a(0) = a_0$, $b(0) = b_0$, and $c(0) = c_0$. The three variables are stoichiometrically related by two conservation relations derived from equation (4.33a), which can be used to eliminate two variables, $b(t)$ and $c(t)$ for example, yielding the remaining single degree of freedom as $\frac{\mathrm{d}a}{\mathrm{d}t} = \frac{\mathrm{d}b}{\mathrm{d}t} = -\frac{\mathrm{d}c}{\mathrm{d}t}$ corresponding to $R = 1$ (see figure 4.5):

$$\begin{aligned} a(t) \,+\, c(t) \;&=\; a_0 \,+\, c_0 \;=\; \vartheta_0^{(ac)} \;, \\ b(t) \,+\, c(t) \;&=\; b_0 \,+\, c_0 \;=\; \vartheta_0^{(bc)} \;, \quad \text{and} \\ b(t) \,-\, a(t) \;&=\; b_0 \,-\, a_0 \;=\; \vartheta_0^{(b)} \;. \end{aligned}$$

One out of these three conditions is dependent, since the second line minus the first line yields the third line. Eventually one finds:

$$\frac{\mathrm{d}a}{\mathrm{d}t} \;=\; -\,k\;a\,b \;=\; -\,k\,a\,(\vartheta_0^{(b)} - a) \;. \tag{4.33g}$$

The ODE is solved by standard techniques and we obtain the solutions

$$\begin{aligned} a(t) \;&=\; \frac{a_0\,\vartheta_0^{(b)}\,\exp\left(-\vartheta_0^{(b)}\,kt\right)}{\vartheta_0 \,+\, a_0\left(1 - \exp\left(-\vartheta_0^{(b)}\,kt\right)\right)} \quad \text{for} \;\; \vartheta_0^{(b)} > 0\,,\; b_0 > a_0\,, \\ a(t) \;&=\; \frac{a_0\,|\vartheta_0^{(b)}|}{a_0 \,-\, (a_0 - |\vartheta_0^{(b)}|)\left(1 - \exp\left(-|\vartheta_0^{(b)}|\,kt\right)\right)} \\ &\qquad \text{for} \;\; \vartheta_0^{(b)} < 0\,,\; b_0 < a_0\,, \quad \text{and} \\ a(t) \;&=\; \frac{a_0}{1 \,+\, a_0\,kt} \quad \text{for} \;\; \vartheta_0^{(b)} = 0\,,\; b_0 = a_0 \;. \end{aligned} \tag{4.33h}$$

by direct integration. The three cases differ in the long-time behavior: $\lim_{t\to\infty} a(t) = 0$ for $\vartheta_0^{(b)} \geq 0$, $b_0 > a_0$ and $\lim_{t\to\infty} a(t) = b_0 - a_0$ for $\vartheta_0^{(b)} < 0$, $b_0 > a_0$.

The reversible bimolecular conversion reaction: $\mathbf{A} + \mathbf{B} \to \mathbf{C} + \mathbf{D}$. The second case simply consists of a reversible bimolecular conversion reaction that is decomposed into two elementary reactions of type (4.1i):

$$\mathbf{A} + \mathbf{B} \xrightarrow{k} \mathbf{C} + \mathbf{D} \quad \text{and} \tag{4.34a}$$

$$\mathbf{C} + \mathbf{D} \xrightarrow{l} \mathbf{A} + \mathbf{B} \,. \tag{4.34b}$$

For the three sets of the chemical reaction network we have

$$\mathcal{S} = \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\} \,, \tag{4.34c}$$

$$\mathcal{C} = \{\mathbf{C}_1 = \mathbf{A} + \mathbf{B}, \mathbf{C}_2 = \mathbf{C} + \mathbf{D}\} \,, \quad \text{and} \tag{4.34d}$$

$$\mathcal{R} = \{\mathbf{R}_1 = \mathbf{C}_1 \to \mathbf{C}_2, \mathbf{R}_2 = \mathbf{C}_2 \to \mathbf{C}_1\} \,. \tag{4.34e}$$

The stoichiometric matrix S is of dimension 4×2:

$$\mathrm{S} = \begin{pmatrix} -1 & +1 \\ -1 & +1 \\ +1 & -1 \\ +1 & -1 \end{pmatrix} \,. \tag{4.34f}$$

In deterministic mass action kinetics, $v^{(\mathrm{MA})}$, the variables are the concentrations of the molecular species, $[\mathbf{A}] = a(t)$, $[\mathbf{B}] = b(t)$, $[\mathbf{C}] = c(t)$, and $[\mathbf{D}] = d(t)$. In order to solve the kinetic differential equation we require two rate parameters, k and l, and four initial conditions: $a(0) = a_0$, $b(0) = b_0$, $c(0) = c_0$, and $d(0) = d_0$. The four variables are stoichiometrically related by three conservation relations in (4.34a) and (4.34b)

$$\begin{aligned} a(t) + b(t) + c(t) + d(t) &= a_0 + b_0 + c_0 + d_0 \,, \\ a(t) - b(t) &= a_0 - b_0 \,, \quad \text{and} \\ c(t) - d(t) &= c_0 - d_0 \,, \end{aligned}$$

and only one degree of freedom – corresponding to the rank $R = 1$ of the stoichiometric matrix – remains: $\mathrm{d}a/\,\mathrm{dt} = \mathrm{d}b/\,\mathrm{dt} = -\mathrm{d}c/\,\mathrm{dt} = -\mathrm{d}d/\,\mathrm{dt}$. Accordingly, we can substitute $b(t) = b_0 - a_0 + a(t)$, $c(t) = c_0 + a_0 - a(t)$, and $d(t) = d_0 + a_0 - a(t)$ and the ODE for the last remaining variable $a(t)$ takes on the form:

$$\begin{aligned} \frac{\mathrm{d}a}{\mathrm{dt}} &= -k\,a\,b + l\,c\,d = -k\,a\,(\vartheta_0^{(b)} + a) + l\,(\vartheta_0^{(c)} - a)(\vartheta_0^{(d)} - a) = \\ &= (h-k)\,a^2 - (k\vartheta_0^{(b)} + l\vartheta_0^{(c)} + l\vartheta_0^{(d)})\,a + l\,\vartheta_0^{(c)}\vartheta_0^{(d)} \,, \end{aligned} \tag{4.34g}$$

where the initial conditions are contained in the quantities $\vartheta_0^{(b)} = b_0 - a_0$, $\vartheta_0^{(c)} = c_0 + a_0$, and $\vartheta_0^{(d)} = d_0 + a_0$.

Equation (4.34g) can be integrated by standard methods to yield an implicit solution of the form $t = f(a)$ but the expression is so clumsy that we dispense here from listing it. The analytical solution for the irreversible forward reaction are identical with the solutions of the association reaction (4.33h)

treated in the previous example, since the kinetic ODEs of an irreversible reaction do not depend on the concentrations on the product side. Clearly, the expressions are also valid for the irreversible backward reaction by replacing $a \leftrightarrow c$, $b \leftrightarrow d$, and $k \leftrightarrow l$.

The Feinberg mechanism (figure 4.7). Our third example is taken directly from Feinberg [125, 127] and deals with six elementary reactions involving five chemical species related by the following mechanism:

$$\begin{aligned} \mathbf{A} &\xrightarrow{k_1} 2\,\mathbf{B}\,, \\ 2\,\mathbf{B} &\xrightarrow{l_1} \mathbf{A}\,, \\ \mathbf{A} + \mathbf{C} &\xrightarrow{k_2} \mathbf{D}\,, \\ \mathbf{D} &\xrightarrow{l_2} \mathbf{A} + \mathbf{C}\,, \\ \mathbf{D} &\xrightarrow{k_3} \mathbf{B} + \mathbf{E}, \quad \text{and} \\ \mathbf{B} + \mathbf{E} &\xrightarrow{k_4} \mathbf{A} + \mathbf{C}\,. \end{aligned} \tag{4.35a}$$

The three sets defining the chemical reaction network are:

$$\mathcal{S} = \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}\}\,, \tag{4.35c}$$

$$\mathcal{C} = \{\mathbf{C}_1 = \mathbf{A}, \mathbf{C}_2 = 2\mathbf{B}, \mathbf{C}_3 = \mathbf{A} + \mathbf{C}, \mathbf{C}_4 = \mathbf{D}, \mathbf{C}_5 = \mathbf{B} + \mathbf{E}\}\,, \quad \text{and} \tag{4.35d}$$

$$\begin{aligned} \mathcal{R} = \{&\mathbf{R}_1 = \mathbf{C}_1 \to \mathbf{C}_2, \mathbf{R}_2 = \mathbf{C}_2 \to \mathbf{C}_1, \mathbf{R}_3 = \mathbf{C}_3 \to \mathbf{C}_4, \\ &\mathbf{R}_4 = \mathbf{C}_4 \to \mathbf{C}_3, \mathbf{R}_5 = \mathbf{C}_4 \to \mathbf{C}_5, \mathbf{R}_6 = \mathbf{C}_5 \to \mathbf{C}_3\}\,. \end{aligned} \tag{4.35e}$$

The stoichiometric matrix S for the mechanism (4.35a) is readily obtained:

$$\mathrm{S} = \begin{pmatrix} -1 & +1 & -1 & +1 & +1 & 0 \\ +2 & -2 & 0 & 0 & -1 & +1 \\ 0 & 0 & -1 & +1 & +1 & 0 \\ 0 & 0 & +1 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & +1 \end{pmatrix}, \tag{4.35f}$$

it has the dimension 5×6 and its rank is $R = 3$. The reaction graph corresponding to this mechanism is shown in figure 4.7. The comparison of both graphs is a nice illustration of one already mentioned property of reaction graphs: The graph visualizes only the interconversions between reaction complexes and contains no information about the molecular realization of the kinetic reaction network, whereas the graphical representation of the reaction network in contains the full information except the specific initial conditions. Analytical solutions for the reaction network (4.35a) are not available but

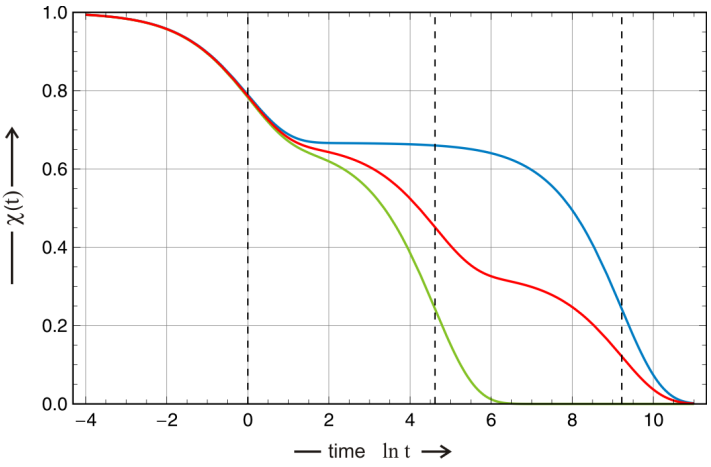


Fig. 4.8 Relaxation times from multi-mode relaxation. Relaxation times can be detected readily as a point of inflection in the plot of $\chi(t)$ against $\ln t$. Multiple relaxation times are found easily when the relaxation processes appear well separated on the time axis. The three curves show two relaxation processes separated by factors 100 (green curve) or 10 000 (blue curve) on the time axis, and three relaxations with relaxation times 1/100/10 000 (red curve). All amplitudes are chosen 1/3 or 2/3 (second process in the green and the blue curve); the time scale is $\ln t$. In cases where the individual processes are not so well separated on the time axis the problem to calculate the relaxation times may be ill-posed [3, p. 252] (see also section 4.1.5).

numerical integration for given initial conditions is easily achieved. Some qualitative properties will be derived in the forthcoming paragraphs.

Multidimensional relaxation. Chemical relaxation theory applied to a single-step reactions was presented in section 4.1.1. It can be readily extended to an arbitrary number of chemical reactions [397]. For K reactions the elements of the relaxation matrix are of the form

$$\mathrm{A} \;=\; \left\{ a_{ij} = -\sum_{k=1}^{K} (\nu_{ki} - \nu'_{ki}) \frac{\nu_{kj}\,(\bar{v}_k)_{\rightarrow} \;-\; \nu'_{kj}(\bar{v}_k)_{\leftarrow}}{\bar{x}_j} \right\} , \tag{4.36}$$

and we expect to find more than one relaxation mode in the approach towards equilibrium corresponding to more than one relaxation time. In vector notation with $\boldsymbol{\chi} = \mathbf{x} - \bar{\mathbf{x}}$ and the thermodynamic equilibrium as reference state as before the relaxation equation is of the form

$$\frac{\mathrm{d}\boldsymbol{\chi}}{\mathrm{dt}} \;=\; \mathrm{A}\,\boldsymbol{\chi} \quad \text{and} \quad \boldsymbol{\chi}(t) \;=\; \exp(\mathrm{A}\,t)\,\boldsymbol{\chi}(0) \;. \tag{4.7'}$$

The formal exponential function of matrix A in equation (4.7') is readily solved by means of an eigenvalue problem. It is illustrative to symmetrize matrix A by means of a similarity transformation

$$\mathrm{D} \;=\; \mathrm{G}\,\mathrm{A}\,\mathrm{G}^{-1} \;\text{ with }\; \mathrm{G} \;=\; \left\{ g_{ij} = \frac{\delta_{ij}}{\sqrt{\bar{x}_i}} \right\} \quad \text{and}$$

$$\mathrm{D} \;=\; \left\{ d_{ij} = -\sum_{k=1}^{K} \frac{(\nu_{ki} - \nu'_{ki})(\nu_{kj} - \nu'_{kj})\,\bar{v}_k}{\sqrt{\bar{x}_i \bar{x}_j}} \right\} \;,$$

since $(\bar{v}_k)_{\rightarrow} = (\bar{v}_k)_{\leftarrow} = \bar{v}_k$ at equilibrium. The matrices D and A have the same eigenvalues and the fact that A can be transformed to a symmetric matrix has the consequence that all its eigenvalues are real, and for numerical diagonalization simple routines can be applied for. For the original matrix A the diagonalization yields

$$\mathit{\Lambda} \;=\; \mathrm{B}^{-1}\,\mathrm{A}\,\mathrm{B}\,,\; \mathrm{A}\,\mathrm{B} \;=\; \mathrm{B}\,\mathit{\Lambda} \;\text{ with }\; \mathit{\Lambda} = \begin{pmatrix} \tau_1^{-1} & 0 & \dots & 0 \\ 0 & \tau_2^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \tau_n^{-1} \end{pmatrix} \;.$$

The diagonal matrix $\mathit{\Lambda}$ contains the eigenvalues and the matrix $\mathrm{B} = \{b_{ij}\}$ collects the eigenvectors:

$$\mathbf{b}_j \;=\; (b_{1j}, \dots, b_{nj})^{\mathrm{t}} \;\text{ with }\; \mathrm{A}\,\mathbf{b}_j \;=\; \lambda_j\,\mathbf{b}_j \;=\; \frac{1}{\tau_j}\,\mathbf{b}_j \;,$$

which fulfil the simple exponential time dependence

$$\mathbf{b}_j(t) \;=\; \mathbf{b}_j(0)\;\exp\!\left(-\frac{t}{\tau_j}\right) \;.$$

Expressed in the original variables χ_j and using $\mathrm{B}^{-1} = \mathrm{H} = \{h_{ij}\}$ for simplicity the results is:

$$\chi_j(t) \;=\; \frac{\sum_{k=1}^{n} b_{jk}\,\beta_k(0)\,\exp\left(-t/\tau_k\right)}{\sum_{i=1}^{n}\sum_{k=1}^{n} b_{ik}\,\beta_k(0)\,\exp\left(-t/\tau_k\right)} \;\text{ with }\; \beta_k(0) \;=\; \sum_{l=1}^{n} h_{kl}\,\chi_l(0) \qquad (4.37)$$

Since the rank R of the reaction network is commonly smaller than the number of chemical species M some of the eigenvalues will vanish: $\lambda = 0$. The corresponding eigenvectors represent then conservation relations. Alternatively, the constraints can be used to reduce the number of ODE's.

Definition of deficiency. First, we repeat the basic definitions of chemical reaction network theory and we point out how the relevant properties can be obtained:

(i) a *linkage class* is a subset of complexes that are linked be reactions and the number of linkage classes is denoted by L,
(ii) a reaction network is *weakly reversible* if and only if a directed arc leads from every complex to every complex of the network,
(iii) the *reaction vectors* combine reactants and products in the stoichiometric way, $\vec{\mathbf{R}} = -\mathcal{C}_\mathrm{R} + \mathcal{C}_\mathrm{P}$, and
(iv) the *rank* of a reaction network, R is the largest linearly independent set that can be found among its reaction vectors.

The linkage classes of a reaction network are obtained straightforwardly: Each complex is displayed exactly once in the sketch of the network, the complexes are joined by introducing the reaction arrows into the sketch, and linkage classes comprise all complexes joined together. The network in figure 4.7, for example, has $L = 2$ linkage classes.

Strong and weak reversibility are directly seen in the reaction graph: In a strongly reversible network all reactions $\mathbf{R} \in \mathcal{R}$ are reversible,

$$(\mathbf{C}_j \to \mathbf{C}_k \in \mathcal{R}) \implies (\mathbf{C}_k \to \mathbf{C}_j \in \mathcal{R}) \;\forall\; (\mathbf{C}_j, \mathbf{C}_k) \in \mathcal{C} \;. \tag{4.38}$$

Weak reversibility relaxes the condition for strong reversibility in the sense that it is sufficient to be able to reach every species from every species by a sequence of reactions. The network in figure 4.7 is weakly reversible.

The rank of a chemical reaction network is defined as

$$R \doteq \mathrm{rank}\{\mathbf{C}_\mathrm{P} - \mathbf{C}_\mathrm{R} \in \mathbb{R}^\mathcal{S} : \mathbf{C}_\mathrm{R} \to \mathbf{C}_\mathrm{P} \in \mathcal{R}\} \;. \tag{4.39}$$

We illustrate by means of a simple example: The six reaction vectors of the network (4.35a),

$$\{2\mathbf{B} - \mathbf{A}, \mathbf{A} - 2\mathbf{B}, \mathbf{D} - (\mathbf{A} + \mathbf{C}), (\mathbf{A} + \mathbf{C}) - \mathbf{D}, (\mathbf{B} + \mathbf{E}) - \mathbf{D}, (\mathbf{A} + \mathbf{C}) - (\mathbf{B} + \mathbf{E})\},$$

can be contracted to the linearly independent subset of dimension three

$$\{2\mathbf{B} - \mathbf{A}, (\mathbf{A} + \mathbf{C}) - \mathbf{D}, (\mathbf{B} + \mathbf{E}) - \mathbf{D}\} \;.$$

Although the network (4.35a) consists of six reactions, only three of them are linearly independent and accordingly it has rank $R = 3$. It is straightforward to see that every reversible reaction consists of two reactions but only one of them can be linearly independent. The determination of the rank R in small systems is properly done by means of the conservation relations but for larger systems a numerical computation of the rank of the stoichiometric matrix S is usually much faster.

The most important quantity of reaction network theory is the *deficiency* of a reaction system, which is defined in the following equation:

$$\text{Deficiency} \quad \delta \doteq N - L - R\,, \tag{4.40}$$

with N being the number of complexes, L the number of linkage classes, and R the number of degrees of freedom or the rank of the reaction kinetics.

The deficiency of a chemical reaction network is a nonnegative quantity [126] and it determines essential features of the reaction system like the existence of unique equilibria and stationary states.

The deficiency zero theorem. The deficiency zero theorem holds for all chemical reaction networks $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ of deficiency zero and makes three statements [126]:

(i) If the network is not weakly reversible then the ODEs for the reaction system $\{\mathcal{S}, \mathcal{C}, \mathcal{R}, \mathcal{K}\}$ with any arbitrary kinetics $\mathcal{K}$ cannot admit a positive equilibrium, i.e., a stationary point in $\mathbb{R}^M_+$,
(ii) if the network is not weakly reversible then the ODEs for the reaction system $\{\mathcal{S}, \mathcal{C}, \mathcal{R}, \mathcal{K}\}$ with any arbitrary kinetics $\mathcal{K}$ cannot admit a cyclic trajectory containing a positive composition, i.e., a point in $\mathbb{R}^M_+$, and
(iii) if the network is weakly reversible (or reversible) then, for any mass action kinetics $v^{(\mathrm{MA})} = \kappa \in \mathbb{R}^R_+$, the ODEs for the mass action system $\{\mathcal{S}, \mathcal{C}, \mathcal{R}, \kappa\}$ have the following properties: Within each positive stoichiometric compatibility class there exists exactly one equilibrium, this equilibrium is asymptotically stable, and there cannot exist a nontrivial cyclic trajectory in $\mathbb{R}^M_+$.

The third property is a highly important extension of equilibrium thermodynamics because existence and uniqueness of a stable equilibrium in the interior of the positive orthant of concentration space is extended from strictly reversible to weakly reversible systems, from closed systems to closed and open systems of deficiency zero. It is worth stressing again that the statements hold for arbitrary finite dimensions of the reaction system irrespectively of the particular choice of rate parameters – provided they are nonnegative.

The deficiency one theorem. The results of the deficiency zero theorem hold for a much wider class of networks than those with deficiency zero. The extension of the range of validity is encapsulated in the *deficiency one theorem*. For the formulation of the theorem it is important to extend the notion of deficiency to individual linkage classes, which are denoted as

$\mathcal{L} = \{\mathbf{L}_1, \mathbf{L}_2, \ldots, \mathbf{L}_L\}$. The number of complexes in linkage class $\mathbf{L}_j$ is denoted by N_j and since a complex can appear only in one linkage class we have $\sum_{j=1}^{L} N_j = N$. The number of independent degrees of freedom of the ODE or the rank of a linkage class $\mathbf{L}_j$ is denoted by R_j,

$$R_j \doteq \text{rank}\{\mathbf{C}_\text{P} - \mathbf{C}_\text{R} \in \mathbb{R}^{\mathcal{S}} : \mathbf{C}_\text{R} \to \mathbf{C}_\text{P} \in \mathcal{R} \wedge \mathcal{C}_\text{R} \in \mathbf{L}_j\}$$

and we define:

$$\text{Deficiency of class } \mathbf{L}_j : \quad \delta_j = N_j - 1 - R_j \,. \tag{4.41}$$

The class deficiency δ_j is a nonnegative integer like δ. The ranks of the subsystems need not be additive but they fulfil $\sum_{j=1}^{L} R_j \geq R$ and this yields for the deficiency of the total network

$$\delta \geq \sum_{j=1}^{L} \delta_j = N - L - \sum_{j=1}^{L} R_j \,. \tag{4.40'}$$

It is illustrative to consider zero deficiency networks because they are precisely those networks that fulfil both of the conditions:

$$\delta_j = 0 \; \forall \; j = 1, 2, \ldots, L \,, \quad \text{and} \quad \delta = \textstyle\sum_{j=1}^{L} \delta_j = 0.$$

Now are in a position to introduce the deficiency one theorem [126].

Let $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ be a reaction network with L linkage classes, let $\delta = N - L - R$ denote the deficiency of the network, $\delta_j = N_j - 1 - R_j$; $j = 1, \ldots, L$ denote the deficiencies of the individual linkage classes, and assume that the two following conditions are fulfilled:

$$\delta_j \leq 1 \; \forall \; j = 1, 2, \ldots, L \,, \quad \text{and} \quad \delta = \textstyle\sum_{j=1}^{L} \delta_j \; (= 0).$$

If the network is weakly reversible, in particular if it is strongly reversible, then for any mass action kinetics $\kappa \in \mathbb{R}_{>0}^{\mathcal{R}}$ the ODEs for the mass action system $\{\mathcal{S}, \mathcal{C}, \mathcal{R}, \kappa\}$ sustains precisely one equilibrium in each positive stoichiometric compatibility class.

Thus, the deficiency one theorem is a powerful tool for the recognition of reaction system lacking multiple stationary states. In later works the existence of multiple stationary states came in focus [78, 128] and these studies make a bridge between applications in chemistry and in biology. We shall come back to reaction systems with multiple steady states and complex dynamics in the next chapter 5.

4.1.4 Theory of reaction rate parameters

Although this monograph aims at analyzing the problems of stochasticity without touching the question where the input parameters for the models come from, we make one exception in case of chemical kinetics. The reason is twofold: (i) In this case we have a well developed theory that allows for tracing chemical kinetics down to first principles from theoretical physics, and (ii) chemical reaction kinetics is at the same time built upon a true wealth of empirical data that provide a unique proving ground for stochastic models. For a comprehensive understanding of chemical reactions the solutions of kinetic differential equations or chemical master equations have to be complemented by detailed knowledge of the processes at the molecular level. In particular, we need the frequencies or probabilities $\pi(t, \mathrm{dt})$ that quantify the event that a reactant molecule or a reaction complex **C** of some reaction **R**, which has been randomly selected at time t, will react to yield products within the next infinitesimal time interval $[t, t+\mathrm{dt}]$. Under two assumptions, (i) spatial homogeneity assumed to be achieved by fast mixing, and (ii) thermal equilibrium, virtually all chemical reactions fulfil the condition

$$\pi(t, \mathrm{dt}) \;=\; \gamma\,\mathrm{dt}\;, \tag{4.42}$$

where γ is the reaction specific, *deterministic* or *probabilistic rate parameter*.[24] If γ is independent of t, π is simply proportional to dt. The two basic conditions (i) and (ii) are fulfilled likewise for chemical reactions in the vapor phase and in dilute aqueous solutions. The rate parameter is a function of the external conditions like temperature and pressure: $\gamma = \gamma(T, p, \ldots)$. In solution, in particular in aqueous solution other external parameters are important like pH, ionic strength, viscosity, etc.

The task to be solved is to conceive a theory that allows for a derivation of rate parameters and there dependence on external parameters like temperature and pressure from first principles of physics and firm empirical data. Such an ambitious undertaking has been successful or is at least promising for some disciplines of physics as well as for chemical kinetics but such a background theory does not exist for most other probabilistic concepts, particularly not for most of biology, for sociology or for economics. The rate parameters of chemical kinetics and their dependence on external parameters can be deduced, in principle, from quantum mechanics. Here we present a brief digression into the molecular theory of reaction rates in order to illustrate how rate parameters originate from a physical background. Although the rigorous approaches were conceived and tested for reactions in dilute gases, they are with some modifications regularly and successfully applied to reaction in solutions. We start by considering two model equations

[24] Almost always the probabilistic rate parameter γ is identical with the conventional deterministic parameter k, and we shall assume that γ can be interchanged with k whenever needed.

for the calculation of rate parameters: (i) the empirical Arrhenius equation conceived in the nineteenth century equation, and (ii) the Eyring equation that is based on quantum mechanics, in particular *transition state theory*. Then we give an overview of calculations of rate parameters from collision theory of chemical reactions and finally a glance on reactive scattering in the *semiclassical* and quantum mechanical approach. In classical mechanics, the motions of particles satisfy Newton's laws, whereas in quantum mechanics, particles are described by the quantum wavefunction, which is a solution of Schrödinger's equation. The full quantum mechanical calculations are consistent but require expensive computer time consuming calculations, and hence they are tractable only in the simplest cases. An approach is called semiclassical if one part of a system is described by quantum mechanics whereas the other is modeled classically. In the semiclassical theory of chemical reactions quantum mechanics is used for the description of molecules and classical trajectories are applied for the description of the reaction. In an advanced form each trajectory is given a quantum phase so that quantum effects such as interference and tunneling can be described using only classical information.

Model equations. Two equations modeling the temperature dependence of rate parameters found widespread application: The empirical Arrhenius equation conceived in the nineteenth century and the Eyring equation being the result of the *transition state theory*.

The Arrhenius equation. For reaction **R** the probabilistic rate parameter and its temperature dependence is given by

$$\gamma(T) \;=\; k(T) \;=\; A\,\exp\left(-\frac{\mathrm{e_a}}{k_\mathrm{B}T}\right)\;. \tag{4.43}$$

where $\mathrm{e_a}$ is the *activation energy*[25] and A is the so-called pre-exponential factor. Equation (4.43) has been proposed for the temperature dependence of the deterministic rate parameter already in 1884 by the Swedish physicist and chemist Svante Arrhenius. It is still used for the evaluation of rate parameters from known temperature dependence of reaction rates, in particular in the form, $\ln k = \ln A - \mathrm{e_a}/(k_\mathrm{B}T)$. The assumption of a temperature independent pre-exponential factor A can be challenged and more flexible equations are known as modified Arrhenius equation:

$$\gamma(T) \;=\; k(T) \;=\; A\left(\frac{T}{T_0}\right)^n \exp\left(-\frac{\mathrm{e_a}}{k_\mathrm{B}T}\right) \quad \text{or} \tag{4.43'}$$

$$\gamma(T) \;=\; k(T) \;=\; A\,\exp\left(-\frac{\mathrm{e_a}}{k_\mathrm{B}(T-T_0)}\right)\;. \tag{4.43"}$$

[25] The activation energy $\mathrm{e_a}$ is given in Joule per molecule and this implies usage of the Boltzmann constant $k_\mathrm{B} = 1.3806488 \times 10^{-23}\,\mathrm{J\,K^{-1}}$. In chemistry it is common to use kilojoule (kJ) per mole instead of molecule, which we indicate by using E_a for the activation energy and then k_B has to be replaced by the gas constant $R = N_\mathrm{L}k_\mathrm{B}$.

wherein T_0 is a reference temperature. The dimensionless exponent n commonly lies in the range $-1 \leq n \leq 1$.

Transition state theory. The theory of the *transition state* dates back to the early applications of quantum mechanics to chemistry. Roughly a decade after the formulation of quantum mechanics by Erwin Schrödinger and Werner Heisenberg the American physicist Henry Eyring [122] proposed a theory of chemical reactions, which allows for the calculation of rate parameters and which is still use more than 65 years after its invention. It provides an alternative to the fully empirical reaction parameters of the Arrhenius equation [270]. The theory deals with reactant molecules, which come together and form an unstable complex called the *transition state* and the reaction progresses further to yield products. The approach can be understood as a kind of semiempirical theory: The reactant molecules as well as the transition state are described by quantum mechanics but the motion along the reaction coordinate ϱ (figure 4.9) is treated by classical mechanics and pure quantum effects like tunneling are not included but can be added (see, e.g., [306]). In order to be activated for the reaction the reaction complex has to be driven up the reaction coordinate[26] ϱ through energy transfer from other molecular degrees of freedom or from the environment until the local maximum called the transition state is reached. Then the reaction complex travels down the product valley and looses energy that is transferred to other degrees of freedom. The transition state is symbolized by double-dagger (‡) and is treated like a molecular entity except one *unstable vibrational mode* understood as the translational motion along the reaction coordinate ϱ. Thermodynamics is applied to calculate the reaction rate parameter for the reaction

$$\mathbf{A} + \mathbf{B} \overset{K^\ddagger}{\rightleftharpoons} [\mathbf{AB}]^\ddagger \xrightarrow{\quad k^\ddagger \quad} \text{products}$$

by making a quasi-equilibrium assumption for the transition state:

$$K^\ddagger = \frac{[\mathbf{AB}^\ddagger]}{[\mathbf{A}] \cdot [\mathbf{B}]} \,. \tag{4.44}$$

The conventional rate parameter is then obtained from $k = k^\ddagger \cdot K^\ddagger$ and what remains, is to find an expression for the rate $k^\ddagger$ with which the transition state is converted into products.

The transition state is considered as a molecular complex with one uncommon degree of freedom consisting of the motion along the reaction coordinate ϱ, which leads to products. All other $3n-7$ degrees of freedom – or $3n-6$ in case of linear geometries – are handled by conventional statistical mechanics and the equilibrium constant for complex formation is of the form

[26] The reaction coordinate is a combination of atomic movements that leads from reactants to products over the lowest conceivable pass on the energy landscape.

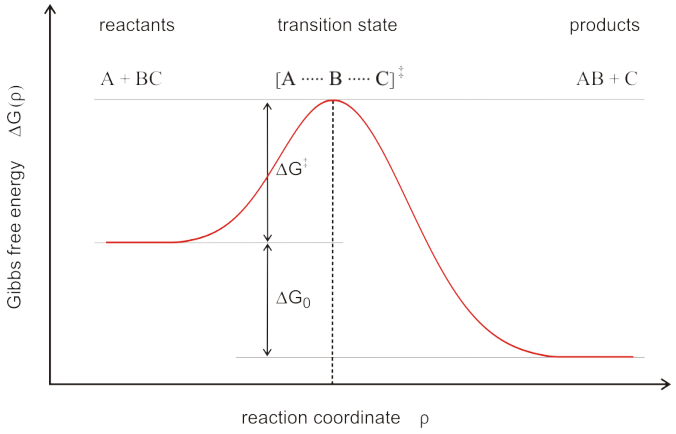


Fig. 4.9 Transition state for the reaction $A + BC \rightarrow AB + C$**.** Reaction dynamics is visualized as a process along a single coordinate called the *reaction coordinate* ϱ. The Gibbs free energy of the reaction complex, $\Delta G(\varrho)$, is plotted against the reaction coordinate and increases during the the approach of the reactants until it reaches a (local) maximum on the energy landscape (see figure 4.12) denoted as transition state. Then through dissipation of free energy to the environment the reaction complex progresses downward in the product valley until it reaches the stable product state. The example presented is an exergonic reaction since $\Delta G_0 = \Delta G_{\text{reactants}} - \Delta G_{\text{products}} < 0$.

$$K^{\ddagger} \;=\; \frac{q_{\mathbf{AB}^{\ddagger}}}{q_{\mathbf{A}}\, q_{\mathbf{B}}}\, e^{-\Delta H_0^{\ddagger}/RT} \;,$$

wherein the individual partition functions are denoted by q and the enthalpy difference between the transition state and the reactants is $\Delta H_0^{\ddagger}$.[27] The remaining degree of freedom is responsible for product formation and has the partition function $q^{(\varrho)}_{\mathbf{AB}^{\ddagger}}$. No matter whether this mode is interpreted as a degenerate vibration with a negative harmonic potential or as a translational degree of freedom we find $k^{\ddagger} \cdot q^{(\varrho)}_{\mathbf{AB}^{\ddagger}} = k_{\mathrm{B}}T/h$ with h being Planck's constant and the final result is the same:

$$k \;=\; k^{\ddagger} \cdot K^{\ddagger} \;=\; \kappa\, \frac{k_{\mathrm{B}}T}{h}\, e^{\Delta S_0^{\ddagger}/R}\, e^{-\Delta H_0^{\ddagger}/RT} \;. \tag{4.45}$$

By κ we denote an empirical transmission factor measuring the probability that the vibrating activated complex decomposes into the product valley, and activation entropy and activation enthalpy are related to the equilibrium

[27] At constant pressure, for example in solution where the volume change ΔV_0 of a reaction is small, the reaction enthalpy ΔH_0 takes on practically the same values as the reaction energy ΔE_0.

constant through:

$$RT \ln K^{\ddagger} = -\Delta G_0^{\ddagger} = -\Delta H_0^{\ddagger} + T\,\Delta S_0^{\ddagger} .$$

Equation (4.45) is Eyring's formula for the value of the reaction rate parameter that corresponds to the rate probability $\gamma_{(\mathbf{A}+\mathbf{B})}$. The value of the formula is twofold: (i) It shows how reaction rate parameters can be derived from first principles, and (ii) it provides a thermodynamic interpretation of the steric factor ρ by means of an activation entropy $\Delta S_0^{\ddagger}$. Direct calculations of rate constants, however, are highly inaccurate since energy surfaces cannot be obtained with sufficient precision apart from a few special cases like the $H+H_2$ reaction (figure 4.12). It should be mentioned that the simpler Arrhenius approach is often preferred over the application of transition state theory to interpretations of temperature dependencies in mechanisms involving biopolymer molecules [465]. There are many possible pitfalls in cases where the reaction mechanisms of the experimental systems are not known in sufficient detail.

Molecular collisions. Molecules or atoms have to come together before they can react, accordingly molecular collisions play a key role in chemical reactions [67], and we present here a short account on molecular collisions (For an excellent introduction into statistical physics of molecular reactions see, e.g., [36, pp. 803-1018]). A vapor phase reaction mixture is assumed in which the molecules behave according to Maxwell-Boltzmann theory. This theory is based on *classical collisions*, which implies that molecules are obeying the laws of Newtonian mechanics, and further it is assumed that the gas is at thermal equilibrium. It has to be remarked, however, that the application of classical collision theory to molecular details of chemical reactions can be only an illustrative and useful heuristic, because the molecular domain falls into the realm of quantum phenomena and any theory that aims at a derivation of reaction probabilities from really first principles should be built upon a firm quantum mechanical basis (see quantum mechanical reaction dynamics).

Molecules change their motions, their internal states, and their natures in collisions, which are classified as elastic, inelastic or reactive, respectively. In an elastic collision the collision partners exchange linear momentum and kinetic energy, and only the directions as well as the absolute values of the velocities of the collision partners before and after the collision are different (figure 4.11). In an inelastic collision internal energy, rotational and/or vibrational and in exceptional cases also electronic energy, is transferred between the reaction partners. Finally, in a reactive collision a chemical reaction takes place between the reaction partners and the molecular species before and after the collision are different. An often made assumption in collision theory is that the colliding objects have spherical geometry, which is apparently a very crude approximation. Corrections can be made by the consideration of a *geometric factor* or by much more elaborate calculations.

In order to be able to handle the specific properties of individual molecules, it is necessary to distinguish molecular species, e.g. $\mathbf{A}$, and individual molecules A.[28] In the latter case knowledge of the detailed molecular state Λ_A may be required, for example,

$$\mathrm{A}_{\Lambda_\mathrm{A}} \quad \text{with} \quad \Lambda_\mathrm{A} = (N_\mathrm{A}, \Sigma_\mathrm{A}, n_\mathrm{A}, J_\mathrm{A}; m_\mathrm{A}, \mathbf{r}_\mathrm{A}, \mathbf{v}_\mathrm{A}) \tag{4.46}$$

where $(N_\mathrm{A}, \Sigma_\mathrm{A}, n_\mathrm{A}, J_\mathrm{A})$ stands for a complete set of molecular quantum numbers characterizing electronic and spin state $(N_\mathrm{A}, \Sigma_\mathrm{A})$, vibrational state (n_A), and rotational state (J_A) of molecule A. The mass of the molecule is m_A, position $(\mathbf{r}_\mathrm{A})$ and velocity coordinates $(\mathbf{v}_\mathrm{A})$ are commonly measured in a Cartesian labor coordinate system: $\mathbf{r}_\mathrm{A}(t) = (x_\mathrm{A}, y_\mathrm{A}, z_\mathrm{A})$ and $\mathbf{v}_\mathrm{A}(t) = \left(v_x^{(\mathrm{A})}, v_y^{(\mathrm{A})}, v_z^{(\mathrm{A})}\right)$. In the spirit of classical mechanics the position vector is – apart from spontaneous changes in collisions – a linear function of time, $\mathbf{r}(t) = \mathbf{r}_0 + \mathbf{v} \cdot t$, and the velocity is constant, $\mathbf{v} = \mathbf{v}_0$, or in other words the molecules travel on straight lines with constant speed between collisions. On this basis we can easily identify the different classes of bimolecular collisions, $\mathrm{A} + \mathrm{B} \rightarrow$, by means of examples where '′' is used to indicate the state after the collision:

(1) Elastic collisions: $\mathrm{A}_{\Lambda_\mathrm{A}} + \mathrm{B}_{\Lambda_\mathrm{B}} \rightarrow \mathrm{A}_{\Lambda_\mathrm{A}} + \mathrm{B}_{\Lambda_\mathrm{B}}$ with $m_\mathrm{A}\mathbf{v}_\mathrm{A} + m_\mathrm{B}\mathbf{v}_\mathrm{B} = m_\mathrm{A}\mathbf{v}'_\mathrm{A} + m_\mathrm{B}\mathbf{v}'_\mathrm{B}$ and $m_\mathrm{A}|\mathbf{v}_\mathrm{A}|^2 + m_\mathrm{B}|\mathbf{v}_\mathrm{B}|^2 = m_\mathrm{A}|\mathbf{v}'_\mathrm{A}|^2 + m_\mathrm{B}|\mathbf{v}'_\mathrm{B}|^2$ corresponding to conservation of linear momentum and kinetic energy. The set of internal quantum numbers remains unchanged in both molecules,
(2) inelastic collisions: $\mathrm{A}_{\Lambda_\mathrm{A}} + \mathrm{B}_{\Lambda_\mathrm{B}} \rightarrow \mathrm{A}_{\Lambda'_\mathrm{A}} + \mathrm{B}_{\Lambda'_\mathrm{B}}$ where the set of quantum numbers for internal motions has been changed by the collision, and
(3) reactive collisions: $\mathrm{A} + \mathrm{B} \rightarrow \cdots$ where the two molecules undergo a chemical reaction in which the nature of at least one molecule is changed.

The correct description of translational motion in a macroscopic reaction vessel does not require quantum mechanical treatment[29] and hence elastic collisions are just an exercise in Newtonian mechanics. Internal energy of molecules is converted into translational energy in inelastic collisions, and a quantum mechanical approach is needed for detailed modeling. The same is true for reactive collisions in case one is interested in reactions of molecules in specific states, otherwise the reaction can described by a mean reaction probability that averages over a Boltzmann ensemble (for the theory of molecular collisions see, e.g., [67]).

Maxwell-Boltzmann distribution. The two conditions, (i) *perfect mixture* and (ii) *thermal equilibrium*, can now be cast into precise physical meanings. Premise (i), *spatial homogeneity*, requires that the probability of finding the

[28] For molecular species we shall also use the notation $\mathbf{X}_1$ when we refer to reaction networks, for example $\boldsymbol{\Xi} = (\mathbf{X}_1, \mathbf{X}_2, \ldots)$.

[29] The individual energy levels of the translational partition function are so close together that the quantum mechanical summation can be replaced by an integral.

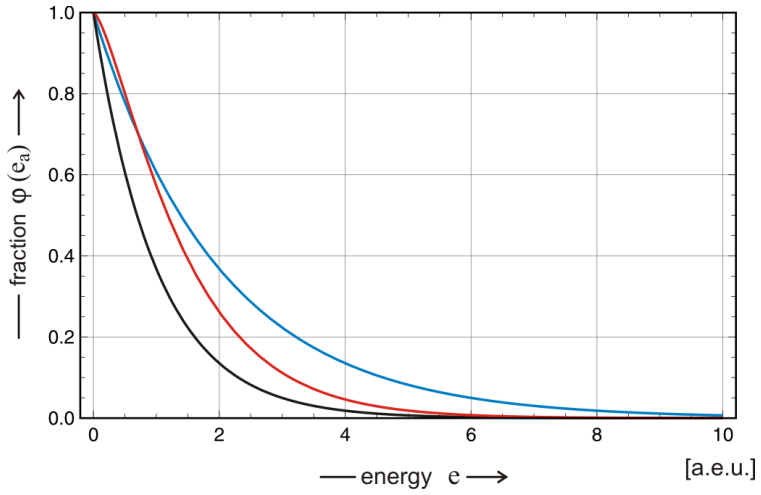


Fig. 4.10 Interpretation of the Arrhenius factor. The fraction of molecules, which have a kinetic energy large than $\mathrm{e_a}$, calculated according to equation (4.49c) (red) is shown together with two simple exponential functions $f_1 = \exp(-\mathrm{e_a}/k_\mathrm{B}T)$ (black) and $f_2 = \exp(-\mathrm{e_a}/(2k_\mathrm{B}T)$ (blue).

center of an arbitrarily chosen molecule inside a container subregion with a volume ΔV is equal to $\Delta V/V$. The system is spatially homogeneous on macroscopic scales but it allows for random fluctuations from homogeneity. Formally, requirement (i) asserts that the position of a randomly selected molecule is described by a random variable, which is uniformly distributed over the interior of the container. Premise (ii), thermal equilibrium, implies that the Cartesian coordinates of the velocity $\mathbf{v} = \big(v_x, v_y, v_z\big)$ with $v = \sqrt{v^2} = \sqrt{v_x^2 + v_y^2 + v_z^2}$ of a randomly chosen particle with mass m are normally distributed with mean $\mu = 0$ and variance $\sigma^2 = k_\mathrm{B}T/m$ (k_B being Boltzmann's constant):

$$f_{MB}(v_i)\,\mathrm{d}v_i \;=\; \left(\frac{m}{2\pi k_\mathrm{B}T}\right)^{1/2} e^{-mv_i^2/(2k_\mathrm{B}T)}\,\mathrm{d}v_i \quad \text{with} \quad i = x, y, z\ . \qquad (4.47)$$

At zero absolute temperature the velocity is a delta-function at $\mathbf{v} = 0$ and it becomes steadily broader with increasing temperature. The extension to the 3D case is straightforward: (i) the velocity densities along the three Cartesian coordinate axes are independent and the expressions are identical by equipartition theorem [37], and (ii) the 3D volume element

$$\mathrm{d}\mathbf{v}^3 \;=\; \mathrm{d}v_x\,\mathrm{d}v_y\,\mathrm{d}v_z \;=\; v^2\mathrm{d}v\,\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \;\;\text{with}\;\; \int_0^{2\pi}\!\!\int_0^{\pi}\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \;=\; 4\pi$$

is evaluated in polar coordinates taking into account spherical symmetry. Then, we obtain the *Maxwell-Boltzmann velocity distribution*

$$\begin{aligned} f_{\mathrm{MB}}(v)\,\mathrm{d}\mathbf{v}^3 \;&=\; \sqrt{\frac{2}{\pi}}\,\frac{v^2}{\alpha^3}\,e^{-v^2/(2a^2)}\,\mathrm{d}\mathbf{v}^3 \;\;\text{and} \\ F_{\mathrm{MB}}(v) \;&=\; \mathrm{erf}\left(\frac{v}{\sqrt{2}\,\alpha}\right) - \sqrt{\frac{2}{\pi}}\,\frac{v}{\alpha}\,e^{-v^2/(2a^2)}\,, \end{aligned} \tag{4.48}$$

where $\alpha = \sqrt{k_{\mathrm{B}}T/m}$. The velocity of molecules is commonly characterized by several averaged values: (i) the mode or the most probable value of the distribution $\widetilde{v}$, (ii) the expectation value $\mathrm{E}(v) = \langle v\rangle$, and (iii) the root mean square velocity derived from the second raw moment, $\left(\widehat{\mu}_2\right)^{1/2} = \sqrt{\langle v^2\rangle}$:

$$\widetilde{v} \;=\; \sqrt{2k_{\mathrm{B}}}\left(\frac{T}{m}\right)^{1/2},\; \langle v\rangle \;=\; \sqrt{\frac{8k_{\mathrm{B}}}{\pi}}\left(\frac{T}{m}\right)^{1/2},\; \sqrt{\langle v^2\rangle} \;=\; \sqrt{3k_{\mathrm{B}}}\left(\frac{T}{m}\right)^{1/2},$$

with $\widetilde{v} < \langle v\rangle < \sqrt{\langle v^2\rangle}$.

It is worth considering the density of the energy as well because is provides a rational explanation for the empirical Arrhenius factor. The total energy is a sum of three independent equal contributions for the three coordinate axes: $\mathrm{e} = \varepsilon_x + \varepsilon_y + \varepsilon_z = 3\,\varepsilon$. In one dimension we find

$$f_{\mathrm{e}1}(\varepsilon)\,\mathrm{d}\varepsilon \;=\; \frac{1}{\sqrt{\pi k_{\mathrm{B}}T}}\,\frac{1}{\sqrt{\varepsilon}}\,e^{-\varepsilon/k_{\mathrm{B}}T}\,\mathrm{d}\varepsilon\;, \tag{4.49a}$$

which by inserting $\varepsilon = x\,k_{\mathrm{B}}T/2$ is easily shown to be a χ^2-density of dimension one:

$$f_{\chi^2_1}(x)\,\mathrm{d}x \;=\; \frac{1}{\sqrt{2\pi}}\,\frac{1}{\sqrt{x}}\,e^{-x/2}\,\mathrm{d}x\;.$$

Extension to three dimensions yields

$$f_{\mathrm{e}}(\mathrm{e})\,\mathrm{de} \;=\; 2\sqrt{\frac{\mathrm{e}}{\pi}}\left(\frac{1}{k_{\mathrm{B}}T}\right)^{3/2}\,e^{-\mathrm{e}/k_{\mathrm{B}}T}\,\mathrm{de}\;, \tag{4.49b}$$

and this expression is equivalent to a χ^2-density of dimension three

$$f_{\chi^2_3}(x)\,\mathrm{d}x \;=\; \frac{1}{\sqrt{2\pi}}\,\sqrt{x}\,e^{-x/2}\,\mathrm{d}x\;.$$

The equivalence to the χ^2-distribution is not surprising since the total energy results from $\mathrm{e} = m\mathbf{v}^2/2 = m(v_x^2 + v_y^2 + v_z^2)/2$, a sum of three squares. Equations (4.49) can be used to calculate the percentage of molecules which

have a kinetic energy that is larger than a given reference level $\mathrm{e_a}$:

$$\varphi(\mathrm{e_a}) \;=\; \int_{\mathrm{e_a}}^{\infty} f_{\mathrm{e}}(\mathrm{e})\,\mathrm{de} \;=\; 1 \,-\, F_{\mathrm{e}}(\mathrm{e_a}) \;, \tag{4.49c}$$

where $F_{\mathrm{e}}(\mathrm{e})$ is the cumulative distribution function associated with the density $f_{\mathrm{e}}(\mathrm{e})$. In figure 4.10 we show that $\phi(\mathrm{e_a})$ is not substantially different from the Arrhenius factor, $\exp(-\mathrm{e_a}/k_{\mathrm{B}}T)$.

We summarize: premises (i) and (ii) assert that the distribution of molecular velocities is isotropic and only a function of mass m and temperature T. Implicitly, the two conditions guarantee also that the molecular position and velocity components are all statistically independent of each other. For practical purposes, we expect the two premises to be valid for any dilute gas system at constant temperature in which *nonreactive* molecular collisions occur much more frequently than *reactive* molecular collisions. The extension to dilute solutions is straightforward [37].

Bimolecular reactive collisions. The theory of molecular collisions in dilute gases is the best developed microscopic model for chemical reactions apart from the quantum mechanical approach. It is well suited for providing a rigorous link between molecular motion an chemical kinetics. The rate parameters of general bimolecular reactions are calculated by means of classical mechanics (figure 4.11) and the Maxwell-Boltzmann distribution.

The occurrence of a bimolecular reaction

$$\mathrm{A} \,+\, \mathrm{B} \;\longrightarrow\; \mathrm{C} \,+\, \ldots \tag{4.50}$$

has to be preceded by an encounter of a molecule A with a molecule B. First we calculate the probability of such a collisional encounter in the reaction volume V. For simplicity molecular species are regarded as spheres with specific masses and radii, m_{A} and r_{A} for A and m_{B} and r_{B} for B, respectively. A collision occurs whenever r_{AB}, the center-to-center distance of the two molecules, becomes as small as the sum of the two radii, $(r_{\mathrm{AB}})_{\min} = r_{\mathrm{A}} + r_{\mathrm{B}}$. The probability that a randomly selected pair of $\mathbf{R}_{\boldsymbol{\mu}}$ reactant molecules – $\boldsymbol{\mu} = (\mathrm{A}, \mathrm{B})$ – at time t will collide within the next infinitesimal time interval $[t, t + \mathrm{dt}]$ is defined by $\mathit{\Pi}_{\boldsymbol{\mu}}(t, \mathrm{dt})$ and the geometry of such a collision is sketched in figure 4.11. The probability density of the relative velocity $\widehat{\mathbf{v}} = \mathbf{v}_{\mathrm{A}} - \mathbf{v}_{\mathrm{B}}$ for the randomly selected pair of reactant molecules, precisely the probability of $\widehat{\mathbf{v}}$ lying within an infinitesimal volume element $\mathrm{d}\widehat{\mathbf{v}}^3$ around $\widehat{\mathbf{v}}$ at time t, is denoted by $f\big(\widehat{\mathbf{v}}(t), \boldsymbol{\mu}\big)$ and obtained from Maxwell-Boltzmann theory (4.48):

$$f\big(\widehat{\mathbf{v}}(t), \boldsymbol{\mu}\big)\;\mathrm{d}\widehat{\mathbf{v}}^3 \;=\; \left(\frac{\widehat{m}}{2\pi\, k_{\mathrm{B}}T}\right)^{3/2} \exp\!\big(-\widehat{m}\widehat{v}^2/(2k_{\mathrm{B}}T)\big)\;\mathrm{d}\widehat{\mathbf{v}}^3 \;.$$

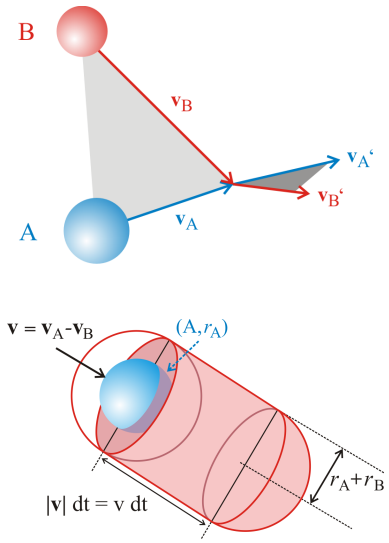


Fig. 4.11 Sketch of molecular collisions in the vapor phase. A *spherical* molecule A with radius r_A moves with a velocity $\mathbf{v} = \mathbf{v}_A - \mathbf{v}_B$ relative to a spherical molecule B with radius r_B. The upper part of the figure shows the geometry of a typical elastic collision, for which linear angular momentum, $\mathbf{p} = m \cdot \mathbf{v}$, and kinetic energy $E_{kin} = m \cdot v^2/2$ are conserved: $\mathbf{p}_A + \mathbf{p}_B = \mathbf{p}'_A + \mathbf{p}'_B$ and $m_A \cdot |\mathbf{v}_A|^2 + m_B \cdot |\mathbf{v}_B|^2 = m_A \cdot |\mathbf{v}'_A|^2 + m_B \cdot |\mathbf{v}'_B|^2$ where the primed quantities refer to the situation after the collision. The lower part of the figure shows the geometry of the collision in the coordinate system of B. If the two molecules are to collide within the next infinitesimal time interval dt, the center of B has to lie inside a cylinder of radius $r = r_A + r_B$ and height $|\mathbf{v}|\,\mathrm{dt} = v\,\mathrm{dt}$. The upper and lower surface of the cylinder are deformed into identically oriented hemispheres of radius r and therefore the volume of the deformed cylinder is identical with that of a non-deformed one.

Herein $\widehat{v} = |\widehat{\mathbf{v}}| = \sqrt{\widehat{v}_x^2 + \widehat{v}_y^2 + \widehat{v}_z^2}$ is the absolute value of the relative velocity and $\widehat{m}$ is the reduced mass of the two molecules A and B.[30]

Next we define a set of all combinations of velocities for the reaction partners in the reaction $\mathbf{R}_{\boldsymbol{\mu}}$ at time t: $\mathcal{R}_{\boldsymbol{\mu}}(t) = \{\mathcal{E}_{\widehat{\mathbf{v}}(t),\boldsymbol{\mu}}\}$. Two properties of the probability densities $f\big(\widehat{\mathbf{v}}(t), \boldsymbol{\mu}\big)$ for different velocities $\widehat{\mathbf{v}}$ are important:
(i) The elements of the set $\mathcal{R}_{\boldsymbol{\mu}}(t)$ of all combinations of velocities of the reactant molecules are mutually exclusive, and

[30] In order to handle relative motion of two particles the original system consisting of particle A with mass m_A and velocity $\mathbf{v}_A$, and particle B with mass m_B and velocity $\mathbf{v}_B$, respectively, is transformed into a system with center of mass (CM) motion and relative or *internal* motion where the center of mass has mass $M = m_A + m_B$ moving with the velocity $\mathbf{v}_{CM} = (m_A\mathbf{v}_A + m_B\mathbf{v}_B)/(m_A + m_B)$ and the internal motion with reduced mass $\widehat{m} = m_A m_B/(m_A + m_B)$ moving with the velocity $\widehat{\mathbf{v}}$.

(ii) they are collectively exhaustive since $\widehat{\mathbf{v}}$ is varied over the entire three dimensional velocity space: $-\infty < (\widehat{v}_x, \widehat{v}_y, \widehat{v}_z) < +\infty$.
The probability density $f\big(\widehat{\mathbf{v}}(t), \boldsymbol{\mu}\big)$ is related to the probability of a collision event $\mathcal{E}_{\text{col}}$ by the conditional probability $P\big(\mathcal{E}_{\text{col}}(t+\,\text{dt})|\mathcal{E}_{\widehat{\mathbf{v}}(t),\boldsymbol{\mu}}\big)$. In figure 4.11 we sketch the geometry of a collision event of two randomly selected spherical molecules A and B, which are assumed to collide within the infinitesimal time interval $[t, t+\,\text{dt}]$:[31] A randomly selected molecule A moves along the vector $\widehat{\mathbf{v}}$ between A and B and a collision of the two molecules will take place in the interval $[t, t+\,\text{dt}]$ if and only if the center of molecule B at time t is situated inside the spherically distorted cylinder indicted in figure 4.11, or the probability of a collision is tantamount to the probability that the center of a randomly selected molecule B is situated within the subregion of V defined by moving A at time t: $P\big(\mathcal{E}_{\text{col}}(t+\,\text{dt})|\mathcal{E}_{\widehat{\mathbf{v}}(t),\boldsymbol{\mu}}\big)$. This subregion has the volume $V_{col} = \widehat{v}\ \sigma_{\text{AB}}\,\text{dt}$, where $\sigma_{\text{AB}} = (r_{\text{A}} + r_{\text{B}})^2\pi$ is the *reaction cross section*, and by scaling with the total volume V we obtain:[32]

$$P\big(\mathcal{E}_{\text{col}}(t+\,\text{dt})|\mathcal{E}_{\widehat{\mathbf{v}}(t),\boldsymbol{\mu}}\big)\ \text{dt} \ = \ \frac{\widehat{v}(t)\cdot\sigma_{\text{AB}}}{V}\ \text{dt}\ . \tag{4.51}$$

The desired probability is calculated through substitution and integration over the entire velocity space

$$\mathit{\Pi}_{\boldsymbol{\mu}}(t,\,\text{dt}) \ = \iiint_{\mathbf{v}=-\infty}^{\infty} \left(\frac{\widehat{m}}{2\pi\, k_{\text{B}}T}\right)^{3/2} e^{-\widehat{m}\widehat{v}^{\,2}/(2k_{\text{B}}T)}\cdot\frac{\widehat{v}(t)\,\text{dt}\cdot\sigma_{\text{AB}}}{V}\ \text{d}\widehat{\mathbf{v}}^3\ .$$

The evaluation of the integral is straightforward and yields

$$\mathit{\Pi}_{\boldsymbol{\mu}}(t,\,\text{dt}) \ = \ \left(\frac{8\,k_{\text{B}}T}{\pi V^2}\right)^{1/2}\frac{\sigma_{\text{AB}}}{\sqrt{\widehat{m}}}\ \text{dt}\ . \tag{4.52}$$

The first factor contains only constants and macroscopic quantities, the volume V and the temperature T, whereas the molecular parameters, the radii r_{A} and r_{B} and the reduced mass $\widehat{m}$ appear in the second factor.

A collision is a necessary but not a sufficient condition for a reaction to take place and therefore we introduce a *collision-conditioned reaction probability* $p_{\boldsymbol{\mu}}$ that is the probability that a randomly selected pair of colliding $\mathbf{R}_{\boldsymbol{\mu}}$ reactant molecules will indeed react according to $\mathbf{R}_{\boldsymbol{\mu}}$. By multiplication of independent probabilities and with respect to equation (4.42) we find

---

[31] The absolute time t comes into play because the positions of the molecules, $\mathbf{r}_{\text{A}}$ and $\mathbf{r}_{\text{B}}$, and their velocities, $\mathbf{v}_{\text{A}}$ and $\mathbf{v}_{\text{B}}$, depend on t.

[32] Implicitly in the derivation we made use of the infinitesimally small size of dt. Only if the distance $\widehat{v}$ dt is vanishingly small, the possibility of collisional interference of a third molecule can be neglected.

$$\pi_{\boldsymbol{\mu}}(t,\,\mathrm{dt}) \;=\; \gamma_{\boldsymbol{\mu}}\,\mathrm{dt} \;=\; p_{\boldsymbol{\mu}}\,\Pi_{\boldsymbol{\mu}}(t,\,\mathrm{dt}) \;=\; p_{\boldsymbol{\mu}}\left(\frac{8\,k_{\mathrm{B}}T}{V}\right)^{1/2}\frac{\sigma_{\mathrm{AB}}}{\sqrt{\widehat{m}}}\,\mathrm{dt}\;. \quad (4.53)$$

As required by equation (4.42) $\gamma_{\boldsymbol{\mu}}$ is independent of dt and this will be the case if and only if reaction probability $p_{\boldsymbol{\mu}}$ does not depend on dt. This is highly plausible for the above given definition, and an illustrative check through the detailed examination of bimolecular reactions can be found in [169, pp. 413-417].

The Arrhenius factor can be illustrated within the frame of collision theory if we make the assumption that the collision energy has to exceed the activation energy $\mathrm{e_a}$. The fraction of molecules whose kinetic energies exceed this energy threshold, $\phi(\mathrm{e_a})$, is readily calculated from the energy distribution function (4.49b) as shown in equation (4.49c). In figure 4.10 $\phi(\mathrm{e_a})$ is compared with the conventional Arrhenius factor $\exp(-\mathrm{e_a}/k_{\mathrm{B}}T)$ and the factor $\exp(-\mathrm{e_a}/2k_{\mathrm{B}}T)$. The second case is rationalized by the idea that both reaction partner contribute an equal share, $\mathrm{e_a}/2$ to the reaction energy. Although there are recognizable differences between the three curves in the figure, the entirely empirical Arrhenius equation parallels nicely the factor $\phi(\mathrm{e_a})$ derived from collision theory.

The results of collision theory for reactive bimolecular encounters can be summarized in a commonly used form for the rate parameter and its temperature dependence

$$\gamma_{\boldsymbol{\mu}}(T) \;=\; A\left(\frac{T}{T_0}\right)^{n}\exp\left(-\frac{\mathrm{e_a}}{k_{\mathrm{B}}T}\right) \;=\; \zeta(T)\,\rho\,\exp\left(-\frac{\mathrm{e_a}}{k_{\mathrm{B}}T}\right)\;. \quad (4.43')$$

Herein $\zeta(T)$ is the collision frequency as calculated above

$$\zeta(T) \;=\; \sigma_{\mathrm{AB}}\sqrt{\frac{8\,k_{\mathrm{B}}T}{\pi\widehat{m}}} \quad \text{with} \quad \sigma_{\mathrm{AB}} = (r_{\mathrm{A}} + r_{\mathrm{B}})^2\,\pi. \quad (4.54)$$

The factor ρ is a denoted as steric factor and $\mathrm{e_a}$ is called the activation energy of the reaction that is measured here as energy per molecule. Often concentrations instead of particle numbers are used and this implies multiplication by Avogadro's number. Then the activation energy, $E_{\mathrm{a}} = N_{\mathrm{L}}\,\mathrm{e_a}$, is commonly given in $[\mathrm{kJ{\cdot}mol^{-1}}]$ and the gas constant $R = N_{\mathrm{L}}\,k_{\mathrm{B}}$ is used instead of Boltzmann's constant. The actual number of collisions in the volume V per time unit is $Z = N_{\mathrm{L}}\,V\,\zeta$. Comparison with the Arrhenius equation (4.43') yields $n = 1/2$. The exponential temperature dependence of the rate parameter on temperature is often fulfilled with astonishingly high accuracy but an interpretation of the steric factor ρ is often unsatisfactory and therefore some chemists prefer to stay away from any rationalization of the steric factor and define it simply as the ratio between the pre-exponential factor and the collision frequency: $\rho = A/\zeta(T)$.

Monomolecular reactions. A *monomolecular reaction* in the strict sense describes the spontaneous conversion

$$\mathbf{A} \longrightarrow \mathbf{C}\,. \tag{4.55}$$

One molecule A is converted spontaneously into one molecule C. The monomolecular reaction has been initially considered as being particularly simple, because only one type of molecule is involved, but this expectation turned out to be wrong: Most formally monomolecular reactions follow a bimolecular rate law at least at sufficiently low concentrations and have to be distinguished from true monomolecular conversions. It is worth mentioning that even a class of spontaneous dissociation reactions of small cluster ions, for example $(H_3O^+)(H_2O)_n$ or $Cl^-(H_2O)_n$ with n=2-4, which were considered as the prototypes of truly monomolecular processes are not strictly spontaneous, because the loss of ligands seems to be initiated by collisions with the wall of the reaction vessel [369].

In absence of interaction with an environment the true monomolecular conversion (4.55) is necessarily driven by some quantum mechanical mechanism similar to the radioactive decay of a nucleus. Time-dependent perturbation theory in quantum mechanics [327, pp. 724-739] shows that almost all weakly perturbed energy-conserving transitions have linear probabilities of occurrence in time intervals δt, when δt is *microscopically large* but *macroscopically small*. Therefore, to a good approximation the probability for a radioactive nucleus to decay within the next infinitesimal time interval dt is of the form $\alpha\,\mathrm{dt}$, were α is some time-independent constant. On the basis of analogy we may expect $\pi_{\boldsymbol{\mu}}(t,\,\mathrm{dt})$ the probability for a genuine monomolecular conversion to be approximately of the form $\gamma_{\boldsymbol{\mu}}\,\mathrm{dt}$ with $\gamma_{\boldsymbol{\mu}}$ being independent of dt.

The vast majority of apparently monomolecular reactions, however, follow a different mechanisms and involve a reaction partner in the sense of a catalyzed bimolecular conversion

$$\mathbf{A} + \mathbf{B} \longrightarrow \mathbf{C} + \mathbf{B} \text{ or} \tag{4.50'}$$

$$\mathbf{A} + \mathbf{A} \longrightarrow \mathbf{C} + \mathbf{A}\,. \tag{4.50"}$$

In equation (4.50') the conversion $\mathbf{A} \longrightarrow \mathbf{C}$ is initiated by a collision of a molecule A with a molecule B, which acts as a catalyst, since it is not consumed by the process.[33] When the collision partner is another molecule A (4.50"), we are dealing with a monomolecular reaction in the conventional sense, which is described straightforwardly as a special class of bimolecular processes.

[33] Formally we are dealing with a reaction that is catalyzed by a molecule of the same or another molecular species and the reaction is related to the spontaneous conversion by rigorous thermodynamics: Whenever a catalyzed reaction appears in a mechanism the uncatalyzed process has to be considered as well, no matter how slow it is.

The first proposal of mechanism (4.50") for the monomolecular conversion was made already in 1922 by Frederick Lindemann [291]: He suggested that the monomolecular conversion is a two step mechanism of the form

$$\mathbf{A} + \mathbf{A} \underset{l_1}{\overset{k_1}{\rightleftharpoons}} \mathbf{A} + \mathbf{A}^* \quad \text{and} \quad \mathbf{A}^* \xrightarrow{k_2} \mathbf{C} \tag{4.56}$$

with $k_2 \ll l_1$. The Lindemann mechanism with a conventional rate parameter k_1 did not fit the experimental data and has been improved by Cyril Hinshelwood [208] by a different interpretation of the activation of molecule A that was extended to a range of energy values $k_{1(E_0 \to E_1)} \Rightarrow k_{1(E_0 \to E_1 + \delta E)}$. Later on the molecular mechanistic details were improved and the Lindemann-Hinshelwood mechanism has been substantially extended by Oscar Rice, Herman Ramsperger [380], and Louis Kassel [240] through the explicit introduction of a transition state $\mathbf{A}^\ddagger$:

$$\begin{aligned} &\mathbf{A} + \mathbf{A} \underset{h_1}{\overset{k_1}{\rightleftharpoons}} \mathbf{A} + \mathbf{A}^* \quad \text{and} \\ &\mathbf{A}^* \xrightarrow{k_{2a}} \mathbf{A}^\ddagger \xrightarrow{k^\ddagger} \mathbf{C}\,. \end{aligned} \tag{4.57}$$

As in transition state theory the rate parameter $k^\ddagger$ corresponds to the fast process associated with the reactive mode of the transition state. Since $k^\ddagger$ is thought to be larger than any other rate parameter, the rate limiting step of the formation of the product $\mathbf{C}$ is the conversion $\mathbf{A}^* \to \mathbf{A}^\ddagger$ and comparing Lindemann and RRK mechanism we have $k_2 \approx k_{2a}$ and $k_{2a} = k^\ddagger[\mathbf{A}^\ddagger]/[\mathbf{A}^*]$ from the steady state assumption. Eventually, the theory of monomolecular reactions got its present form through a reformulation of the transitions state by Rudolph Marcus and Oscar Rice [305, 307, 308]. The current version of the so-called RRKM theory of monomolecular reactions theory allows for a highly accurate and very detailed description of reactions and it can be readily converted into a stochastic model [288].

Termolecular and other reactions. Termolecular or trimolecular reactions of the form

$$\mathbf{A} + \mathbf{B} + \mathbf{C} \longrightarrow \mathbf{D} + \ldots \tag{4.58}$$

are rare and need not be considered, because collisions of three particles do not occur with a probability larger than of measure zero. Exceptions are two classes of reactions: (i) vapor phase association reactions where a third body is required as collision partner removing energy and (ii) the reaction of nitrogen monoxide with oxygen or halogens. A characteristic example of a class (i) reaction is the formation of ozone

$$\mathbf{O} + \mathbf{O}_2 + \mathbf{N}_2 \longrightarrow \mathbf{O}_3 + \mathbf{N}_2\,,$$

where the nitrogen molecule removes energy in order to allow for reaching a bound state of ozone [359]. The typical class (ii) reaction is the oxidation of nitrogen oxide with molecular oxygen [353]

$$2\,\mathbf{NO} + \mathbf{O}_2 \longrightarrow 2\,\mathbf{NO}_2 \,.$$

Although nitric oxide oxidation by oxygen is considered as the prototype of a termolecular reaction two competitive two step mechanism involving only bimolecular collisions are discussed as well:

$$\mathbf{NO} + \mathbf{NO} \rightleftharpoons (\mathbf{NO})_2\,,\ (\mathbf{NO})_2 + \mathbf{O}_2 \rightarrow 2\,\mathbf{NO}_2 \ \text{ or}$$
$$\mathbf{NO} + \mathbf{O}_2 \rightleftharpoons \mathbf{NO}_3\,,\ \mathbf{NO}_3 + \mathbf{NO} \rightarrow 2\,\mathbf{NO}_2 \,.$$

A comparison of the data for all three mechanistic variants of the reaction are found in the review [431]. However, there may also be, special situations where approximations of complicated processes by termolecular events is justified. One example is a set of three coupled reactions with four reactant molecules [168, pp. 359-361] where $\pi_{\boldsymbol{\mu}}(t,\,\mathrm{dt})$ is essentially linear in dt.

Zeromolecular reactions. The last class of reaction to be considered here is no proper chemical reaction but ,for example, an inflow of material into the reactor. It is often denoted as a the zeroth order reaction (4.1a):

$$* \longrightarrow \mathbf{A}\,. \tag{4.59}$$

Here, the assumption that the inflow is accompanied by *efficient mixing* fulfilling the homogeneity condition is essential, because it guarantees that the number of molecules entering the homogeneous system per time unit is a constant, and does not depend on dt.

Quantum mechanical reaction dynamics. For any detailed understanding of chemical reactions from first physical principles knowledge from quantum mechanics is indispensable. Here we direct readers to the great variety of existing text books (recently published monographs are [23, 396] for basic quantum mechanics and applications in chemistry, and a quite elaborate text on dynamics is found in [310]) and sketch only the basic idea because of its general importance: In conventional quantum chemistry the fast motions of electrons are separated from the slow motions of atomic nuclei and the stationary Schrödinger equation of a molecule or a reaction complex is partitioned into two equations

$$\mathcal{H}_{\mathrm{el}}\,\Psi_{\mathrm{el}}^{(n)}(\mathbf{r}) \;=\; E_n(\mathbf{R})\,\Psi_{\mathrm{el}}^{(n)}(\mathbf{r}) \;\text{ with }\; \mathcal{H}_{\mathrm{el}} \;=\; \mathcal{T}_{\mathrm{el}} + V(\mathbf{r},\mathbf{R})\,, \tag{4.60a}$$
$$\Big(\mathcal{T}_{\mathrm{nuc}} \,+\, E_n(\mathbf{R})\Big)\,\Xi_{\mathrm{nuc}}^{(k;\,n)}(\mathbf{R}) \;=\; W_{k,n}\,\Xi_{\mathrm{nuc}}^{(k;\,n)}(\mathbf{R})\,. \tag{4.60b}$$

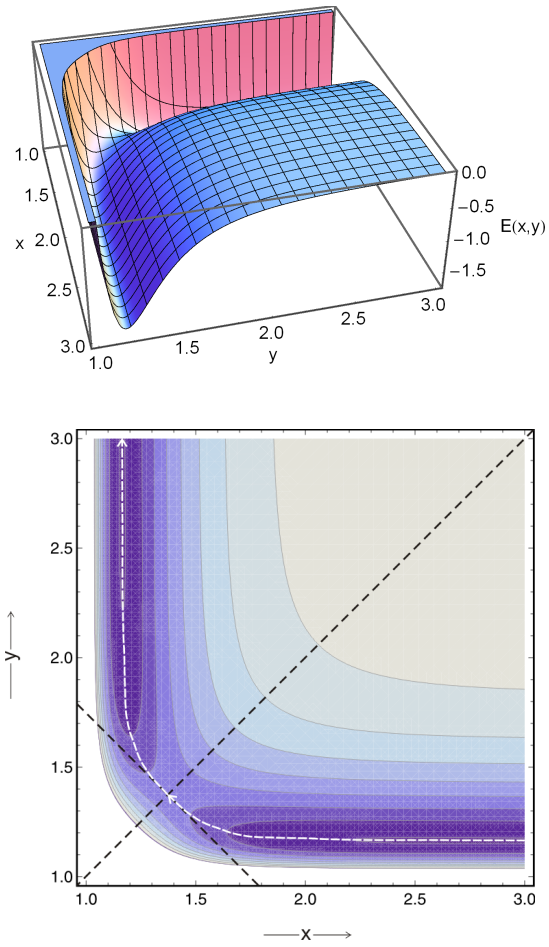


Fig. 4.12 Energy surface of the symmetric bimolecular triatomic exchange reaction A + BC → AB + C**.** The best studied example of such a reaction is the hydrogen isotope exchange reaction D + HD → DH + D for which a highly accurate energy surface is available. The three atoms lie on a straight line. The model surface plotted here is

$$E(x,y) = a/x^{12} - b/x^6 + a/y^{12} - b/y^6 + c/(x+y)^{12}.$$

The upper part of the figure shows a 3D-plot of the energy surface with the reaction path being recognizable as a steep valley. The lower part presents a contour plot of this surface. The broken white line indicates the reaction coordinate ϱ: In the steep horizontal valley at the bottom of the figure the atom D is approaching the molecule HD, then the bond becomes longer and at the saddle point the two bonds are of equal length. Parameters: $a = 10$, $b = 8$, and $c = 1.5 \times 10^5$, leading to a bond length of $r_e = 1.165$ [l.u.] and a bond energy of $\Delta E = -1.6$ [e.u.]. At the saddle point the distance is $x = y = 1.3856$ [l.u.] and the energy amounts to $\Delta E = -1.1303$ [e.u.]. Length and energy are given in arbitrary units, [l.u.] stands for length unit and [e.u.] for energy unit respectively.

The positions of all electrons are subsumed in the vector $\mathbf{r}$, and likewise the nuclei occupy positions denoted by $\mathbf{R}$. Both equations are partial differential equations and they are coupled through the energy hypersurface $E_n(\mathbf{R})$ (see figure 4.12). The Hamilton operator $\mathcal{H}_{\mathrm{el}}$ describes the motion of electrons and consists of the kinetic energy operator of electrons $\mathcal{T}_{\mathrm{el}}$ and the electrostatic potential $V(\mathbf{r},\mathbf{R})$ caused by the electric charges of electrons and nuclei, $E_n(\mathbf{R})$ is the n-th eigenvalue of the Schrödinger equation (4.60a), and $\Psi_{\mathrm{el}}^{(n)}$ is the corresponding eigenfunction. The separation of electronic and nuclear motion was introduced into quantum mechanics by Max Born and Robert Oppenheimer in 1927 [48]. Because of the large difference in mass between electrons and nuclei – being at least three orders of magnitude – and the reasonable assumption that linear momenta of electrons and nuclei are roughly the same because the forces acting on them are identical – *actio equals reactio* – we have

$$M\,\frac{\mathrm{d}\mathbf{R}}{\mathrm{dt}} \;=\; \mathbf{P} \;\approx\; \mathbf{p} \;=\; m\,\frac{\mathrm{d}\mathbf{r}}{\mathrm{dt}} \quad \text{with} \quad M \gg m \quad \text{and hence} \quad \frac{\mathrm{d}\mathbf{R}}{\mathrm{dt}} \ll \frac{\mathrm{d}\mathbf{r}}{\mathrm{dt}}\;.$$

Seen from the fast moving electrons nuclei are practically immobile, the total wave function can be factorized, $\Phi(\mathbf{r},\mathbf{R}) = \Psi_{\mathrm{el}}^{(n)}(\mathbf{r}) \cdot \Xi_{\mathrm{nuc}}^{(k;\,n)}(\mathbf{R})$ or, in other words, the electrons see the nuclei at fixed positions and the nuclei see the electrons in form of a potential coming from a time-averaged mean density. Within the Born-Oppenheimer approximation the connecting piece between the electron density in the quantum state n and nuclear motion but also chemical reactions is the energy hypersurface $E_n(\mathbf{R})$. Classical collision theory (see 'bimolecular reactive collisions') did not explicitly account for energetic aspects of reactions and the consideration of an energy surface is an appropriate and important extension. Nuclear motion can be modeled by Newtonian mechanics and the combination of an energy surface of quantum mechanical origin and classical dynamics is often addressed as semiclassical collision theory in contrast to the full quantum mechanical approach based on scattering theory [67].

Despite the spectacular progress in numerical quantum chemistry many chemical reaction systems and most biologically relevant structures are to large for systematic computational studies, which frequently have to handle the motions of up to 100 000 atoms on time scales of tens of nanoseconds. Hybrid methods combining of quantum mechanical calculations with and molecular mechanics simulations based on Newtonian mechanics (QM/MM) seem to be most promising at present [287, 402, 403].

4.1.5 Empirical rate parameters

The rate parameters – often called rate constants despite the fact that they are no constants in reality and their dependence on external quantities like

temperature, pressure, pH, ionic strengths and other quantities provides insights into reaction mechanisms – are the first quantities derived from measured data, and as such they make sense only for a given mechanism. Very often the mechanism of reaction is not precisely known and then we are confronted with the difficult task to determine the reaction mechanism and the rate parameters simultaneously. Three different approaches are in common use: (i) traditional parameter fitting by means of linearized functions of time dependencies of signals, (ii) parameter fitting by means of computer assisted minimization of a cost function commonly adapted for a given mechanism, and (iii) the mathematically and computationally more expensive but professional method to treat parameter fitting as an inverse problem, which because of its ill-posedness requires regularization in the search for a solution.

The traditional parameter fitting approach was done by hand and we mention the analysis of first order reactions and binding equilibria as a characteristic examples. At present more elaborate methods of parameters evaluation replace the human eye by conventional statistics though employing mean least square fits.

Superposition of exponential curves. First order reactions and in particular all relaxation processes (4.7) follow an exponential function

$$\frac{\mathrm{d}\chi}{\mathrm{dt}} = \frac{1}{\tau_R}\,\chi\,, \quad \chi(t) = \chi_0\,\exp(-t/\tau_R) \quad \text{and} \quad \ln\chi(t) = \log\chi_0 - \frac{t}{\tau_R}\,, \tag{4.61}$$

which plotted on semilogarithmic paper yields a straight line.[34]

In mathematical fitting methods a large number of noisy experimental data points is used to determine a few parameters in the sense of a massively overdetermined problem. Several standard techniques like, for example the method of least squares [43, 466], are available for fitting data to a linear relation. Historically the first documented usage of the least square method is due to the French mathematician Adrien-Marie Legendre who published it 1805 in a monograph [278]. Carl Friedrich Gauß, however, claimed priority in 1895 by contending that he had used the method already 1795.

In linear regression the dependent variables y_i corresponding to measured data are given by

$$y_j = \sum_{i=1}^{m} \beta_i\, x_{j\,i} + \eta_j\,;\; j = 1, \ldots, n\,, \tag{4.62}$$

where n is the number of measured data and m is the number of independent parameters β_i. In vector notation equation (4.62) takes on the form

$$\mathbf{y} = \mathrm{X}\,\boldsymbol{\beta} + \boldsymbol{\eta}\,. \tag{4.62'}$$

[34] We remark that the plot shown in figure 4.8, which was used there to detect several relaxation processes on different time scales, was a $\chi(t)/\log t$-plot whereas the semilogarithmic plot used here is a $\log\chi(t)/t$-plot.

The so-called rectangular *design matrix*,

$$\mathrm{X} \;=\; \{x_{j\,i}\} \;=\; \begin{pmatrix} \mathbf{x}_1^{\mathrm{t}} \\ \mathbf{x}_2^{\mathrm{t}} \\ \vdots \\ \mathbf{x}_m^{\mathrm{t}} \end{pmatrix} \;=\; \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix} \quad \text{with} \quad \mathbf{x}_j \;=\; \begin{pmatrix} x_{j\,1} \\ x_{j\,2} \\ \vdots \\ x_{j\,n} \end{pmatrix} .$$

is multiplied by the parameter vector $\boldsymbol{\beta}$ and after addition of noise $\boldsymbol{\eta}$ we obtain the vector of noisy data $\mathbf{y}$. The basis of the regression is a model function $f(\mathbf{x}, \boldsymbol{\beta})$ that defines the relation between the dependent variables y, the independent variables x, and the parameters $\boldsymbol{\beta}$. The noise term $\boldsymbol{\eta}$ eventually covers all errors in the measured y_j values. Commonly, the errors are assumed to be independent and normally distributed, $f(\eta_j) = \mathcal{N}(\eta_j; 0, \sigma_j^2)$. We distinguish homoscedastic random variables with identical variances, $\sigma_j^2 = \sigma^2 \,\forall\, j$, and the heteroscedastic case with different variances.

In the least squares fitting method the overdetermined equations are solved by means of minimizing a cost function,

$$S \;=\; \sum_{j=1}^{m} r_j^2 \quad \text{with} \quad r_j \;=\; y_j \;-\; f(\mathbf{x}_j, \boldsymbol{\beta}) \;, \tag{4.63}$$

which is given the sum of the squares of the residuals r_j, which by comparison with equation (4.62) are tantamount to the noise. In formal terms we express the minimization by $\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta} \in \mathcal{B}} S(\boldsymbol{\beta})$ where $\mathcal{B}$ is the entire parameter space, and $\widehat{\boldsymbol{\beta}}$ represents the *best* choice of parameters within the frame of the least sum of squares residuals. Linear regression analysis allows for a computation of the best fit in closed form and in a single step whereas iterative methods are commonly required in nonlinear problems. We remark that the term linear refers to the dependence of the dependent variable y on the parameters: The function f need not be linear in the argument x but only in the parameters β_j and, for example, polynomial functions $f(\mathbf{x}_j, \boldsymbol{\beta}) = \sum_i \beta_i x_{j\,i} = \sum_i \beta_i (z_j)^i = y_j - \eta_j$ can be used where we introduced the nonlinear function by setting $x_{j\,i} = (z_j)^i$.

In the simplest method called *ordinary least squares* the error terms η_j are assumed to have finite and identical variances – a property often characterized as homoscedastic – and to be uncorrelated with the independent variables $x_{j\,i}$ and among each other, and then a analytical solution of the parameter estimate is available

$$\widehat{\boldsymbol{\beta}} \;=\; (\mathrm{X}^{\mathrm{t}}\mathrm{X})^{-1}\,\mathrm{X}^{\mathrm{t}}\,\mathbf{y} \;=\; \mathrm{X}^{+}\,\mathbf{y} \tag{4.64a}$$

where $\mathrm{X}^{+} = (\mathrm{X}^{\mathrm{t}}\mathrm{X})^{-1}\,\mathrm{X}^{\mathrm{t}}$, the so-called Moore-Penrose *pseudoinverse matrix*. The equations (4.64) are called *normal equations*, and they are the starting point for the development of numerical techniques for solving linear regression problems.

If the assumptions of uncorrelatedness and homoscedasticity of noise are relaxed the regression problem called *generalized least squares* can be readily handled by lines algebra provided the covariance matrix of the noise terms, $\boldsymbol{\Sigma} = \{\Sigma_{ij} = \mathrm{cov}(\eta_i \eta_j)\}$ (section 2.3.4), is known:

$$\widehat{\boldsymbol{\beta}} \;=\; (\mathrm{X}^{\mathrm{t}} \Sigma^{-1} \mathrm{X})^{-1}\, \mathrm{X}^{\mathrm{t}} \Sigma^{-1}\, \mathbf{y}\ . \tag{4.64b}$$

The matrix $\boldsymbol{\Sigma}$ covers indeed both deviations from the idealized *ordinary* case: The diagonal elements, $\Sigma_{jj} = \sigma_j^2$, take care of heteroscedasticity in case of uncorrelatedness, and the off-diagonal elements, $\Sigma_{j\,i}$, cover correlations between the noise terms in different data. Often the assumption that the errors in all measured point have the same normal distribution is not justified and then equation (4.64b) provides a useful tool for heteroscedastic data sets.

In the light of linear regression we reconsider the determination of the relaxation time τ_R of a single exponential. The calculation is straightforward if we replace in equation (4.61) $\log \chi(t_j) \Leftrightarrow y_j$, $t_j \Leftrightarrow x_{j\,1}$, $\log \chi_0 \Leftrightarrow \beta_0$, and $1/\tau_R \Leftrightarrow \beta_1$, and then the residuals take on the form

$$r_j \;=\; y_j \,-\, \beta_0 - \beta_1 t_j\ ,$$

which allow for direct evaluation by equations (4.64a,b). The superposition of exponentials as it happens in case of multiple relaxations (figure 4.8) and in many other cases gives rise to substantial fitting problems. The American computer scientist Forman Sinnickson Acton [3, 4] characterizes the task of fitting two exponentials as a notoriously ill-posed problem when the two relaxation times differ by less than a factor five. We refer to several original papers dealing with this subject [65, 90, 222, 268] in detail. Finally, we mention a method that allows to fit parameters to a continuous spectrum of infinitely many relaxation processes [377].

Linearization of binding equilibria. The Scatchard equation is presented as an example of a nonlinear relation that is transformed exactly to yield a linear plot (figure 4.13), which has been invented for the analysis of binding equilibria of small ligands to macromolecules. The plot is named after George Scatchard, who was a chemist at the Massachusetts Institute of Technology (MIT). Today it is mainly of historical interest but it still has the advantage that the quality of data can be checked easily by visual inspection. In biochemistry the binding equlibrium $\mathbf{A}{+}\mathbf{B}{\rightleftharpoons}\mathbf{C}$ commonly involves ligand $\mathbf{A}$, a macromolecules usually being a protein $\mathbf{B}$, and the association complex $\mathbf{C} \equiv \mathbf{A} \cdot \mathbf{B}$. We distinguish free concentrations $[\mathbf{A}] = a$, $[\mathbf{B}] = b$, and $[\mathbf{C}] = c$, and total or initial concentration of ligand and protein, $a_0 \;=\; a + c$ and $b_0 = b + c$, respectively. The equilibrium constant

$$K_{\mathrm{b}} \;=\; \frac{[\mathbf{C}]}{[\mathbf{A}] \cdot [\mathbf{B}]} \;=\; \frac{c}{a \cdot b} \quad \text{or} \quad K_{\mathrm{d}} \;=\; K \;=\; \frac{a \cdot b}{c}$$

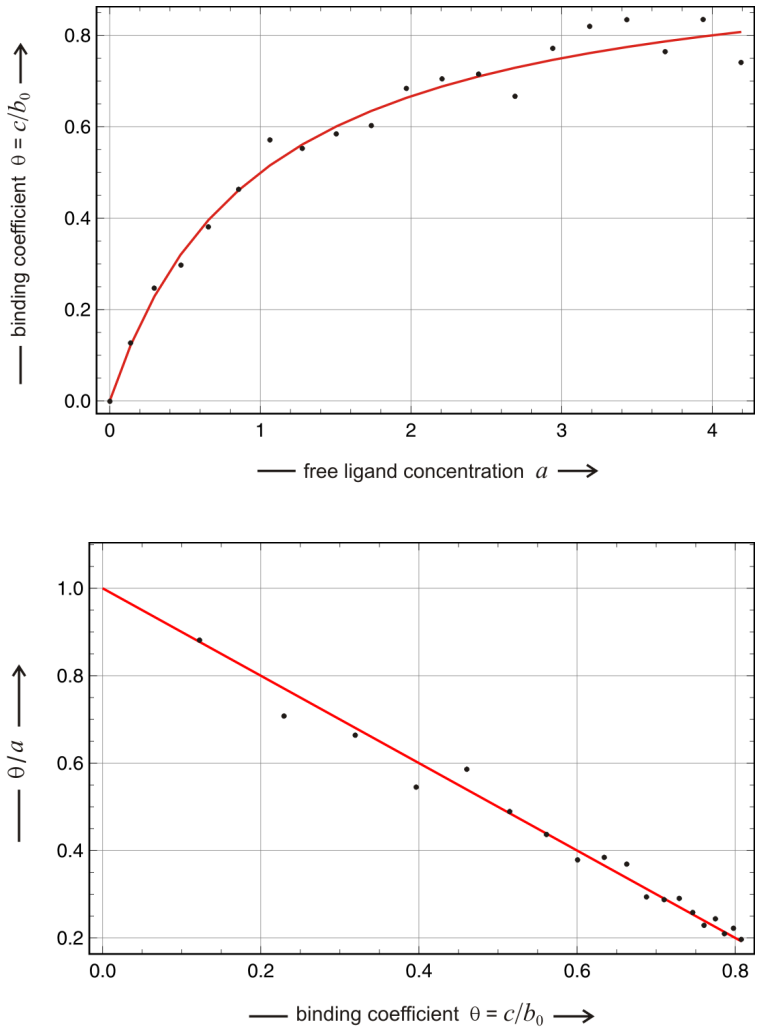


Fig. 4.13 The Scatchard plot and fitting of binding constants. The upper part of the figure shows the hyperbolic binding isotherm of the binding equilibrium **A**+**B**⇌**C** according to equation (4.65): The *degree of saturation* or the *binding coefficient*, $\theta = c/b_0$, is plotted against the free ligand concentration, a. Random scatter is introduced, for example, through errors in the determination of the free concentrations. The lower part presents a Scatchard plot of the same data according to equation (4.66): θ/a is plotted against θ. Parameter choice: $K = 1$, $a_0 = b_0 = 1$.

can be formulated as binding constant K_b or as dissociation constant K_d. Biochemists prefer the latter choice and we shall also adopt it here and drop the subscript: $K_\mathrm{d} = K$. The experimental investigation of binding is conven-

tionally done in terms of the *degree of saturation* or the *binding coefficient* $\theta = c/b_0$ with the property $0 \le \theta \le 1$ where $\theta = 0$ refers to no binding and $\theta = 1$ to complete saturation of the protein. The function $\theta(a)$ is generally characterized as *binding isotherm*[35]

$$\begin{aligned} \theta(a) &= \frac{c}{b_0} = \frac{a}{K + a} \quad \text{with} \\ a &= \frac{1}{2}\left((a_0 - b_0 - K) + \sqrt{4Ka_0 + (a_0 - b_0 - K)^2}\right) , \end{aligned} \tag{4.65}$$

where the free concentration of the ligand, a, is nonlinearly related to the initial concentrations a_0 and b_0. The nonlinear relation (4.65) is not suitable to determine the asymptotic maximum of the binding curve by visual inspection and to adjust a value to the dissociation constant K (figure 4.13) but it can be transformed and one of the resulting linear relations is known as Scatchard plot:

$$\frac{\theta}{a} = \frac{1}{K}(1 - \theta) = \frac{1}{K} - \frac{1}{K}\theta . \tag{4.66}$$

The binding constant is obtained from the slope of the straight line, $\alpha = K^{-1}$ in the plot and in figure 4.13 we show a typical example. The scatter of points was obtained by superimposing a random component on the concentration of the complex **C**. A derivation of the Scatchard equation starts by dividing both sides of equation (4.65)by a: $\theta/a = 1/(K + a)$ and proceeds as follows:

$$\begin{aligned} \frac{1}{K+a} &= \frac{1}{K} - \frac{1}{K} + \frac{1}{K+a} = \frac{1}{K} - \frac{K+a-a}{K(K+a)} = \\ &= \frac{1}{K} - \frac{a}{K(K+a)} = \frac{1}{K} - \frac{\theta}{K} . \end{aligned}$$

□

The $(\theta, \theta/K)$-plot is linear and linear regression can be applied to the binding problem. It is worth that the result is exact and we did not perform a linearization as an approximation.

For current methods in parameter estimation we refer to two monographs as examples of an enormous literature [341, 437]. Present day numerical analysis of measured data is mostly based on the application of inverse methods. We give here a few references to reviews and monographs [21, 113, 114, 425] and mention one a recent paper on parameter analysis of the multistep reaction of chlorite with iodide that aims at the determination of the data sensitive parameters by means of sparsity regularization [266].

[35] The notion *isotherm* points at the fact that the curve is recorded at constant temperature indicating thereby the existence of a pronounced temperature dependence of the equilibrium parameter.

4.2 Stochasticity in chemical reactions

There are two frequently applied techniques for analyzing stochasticity in chemical reaction systems:

(i) modeling and simulation of sample trajectories that correspond to single experimental recordings, or
(ii) analysis of chemical master equations that provide probability distribution over the admissible states as functions of time.

Modeling of sample trajectories through the search for solutions of chemical Langevin equations (section 3.185) is not particularly popular but provides a powerful technique for the analysis of chemical reactions. Simulations of kinetic trajectories (section 4.6) is the computer based counterpart of recording experiments. The simulations involve relatively easy mostly time consuming numerical computations, which provide the desired results but in general they are not very insightful. Solving chemical master equations (section 4.2.1) would be the method of choice were there not the lack of general methods for the solution of nonlinear partial differential equations. The probability densities once derived provide direct access to all information on the chemical processes.

Chemical Langevin and Fokker-Planck equations are based on continuous stochastic variables that are the counterparts of the concentrations in the deterministic equations. Chemical master equations and numerical simulations use discrete stochastic variables $\mathcal{N}(t)$, which represent particle numbers. Hence they can take only nonnegative integer values, $n \in \mathbb{N}_{\geq 0}$, and the probabilities are given by $P_n(t) = P\big(\mathcal{N}(t) = n\big)$. If a running index for integers is needed it will be denoted by m.[36] Some conventions and simplifications in the notation are introduced: We shall use the forward equation unless stated differently and assume an infinitely sharp initial density: $P(n, 0|n_0, 0) = \delta_{n,n_0}$ with $n_0 = n(0)$. Then the full notation can be simplified: $P(n, t|n_0, 0) \Rightarrow P_n(t)$. In addition, the notation $P_n(t)$ indicates already that t is a continuous variable whereas n is discrete. The expectation value of the stochastic variable $\mathcal{N}(t)$ will be denoted by

$$E\big(\mathcal{N}(t)\big) = \langle n(t)\rangle = \sum_{n=0}^{\infty} n \cdot P_n(t) \ , \tag{4.67}$$

and its stationary value, provided it exists, will be expressed as

$$\bar{n} = \lim_{t\to\infty} \langle n(t)\rangle \ . \tag{4.68}$$

In may cases but not always the stationary expectation value $\bar{n}$ will be identical with the long time value of the corresponding deterministic variable.

[36] In cases were more running indices are required we shall use n', m', etc.

4.2.1 The chemical master equation

The *chemical master equation* has been shown to be based on a rigorous microscopic concept of chemical reactions in the vapor phase within the frame of classical collision theory [169] provided two general requirements are fulfilled: (i) a homogeneous mixture as it is assumed to exits through well stirring and (ii) thermal equilibrium implying that the velocities of molecules follow a Maxwell-Boltzmann distribution. Daniel Gillespie's approach focusses on chemical reactions rather than molecular species and is well suited to handle reaction networks. In addition the algorithm can be easily implemented for computer simulation. In section 4.6 we shall discuss the Gillespie algorithm together with computer program implementations. Although the numerical approach is straightforward and yields excellent results for specific examples and small population sizes there is, at the same time, need for an analytical approach in order to find answer to general questions that cannot be given by the numerical simulations.

Exact trajectories. The random vector of particle numbers of M different chemical species at time t, $\vec{\mathcal{X}}(t) = \big(\mathcal{X}_1(t), \ldots, \mathcal{X}_M(t)\big)$, is given by $\vec{\mathcal{X}}(t) = \mathbf{n}(t) = \big(\mathrm{n}_1(t), \ldots, \mathrm{n}_M(t)\big)$. Reaction events occur at times t_k with $k \in \mathbb{N}$ and accordingly the number of $\mathbf{X}_i$ molecules at the end of the k-th interval, $[t_{k-1}, t_k = t_{k-1} + \Delta t_k]$, will be given by

$$\mathbf{n}(t_k) \;=\; \mathbf{n}(t_{k-1}) + \mathbf{s} \quad \text{or} \quad \mathrm{n}_i(t_k) \;=\; \mathrm{n}_i(t_{k-1}) + s_i\,;\; i = 1, \ldots, M\,, \tag{4.69}$$

where s_i is the i-th component of the single reaction stoichiometric vector $\mathbf{s} = (s_1, \ldots, s_M)^{\mathrm{t}}$. The fully fledged computation of $\mathcal{X}_i(t)$ provides a single trajectory of the chemical reaction system. The simulation is called *exact* because the probability distribution obtained by sampling of trajectories converges statistically to the corresponding solution of the analogue master equation. In figure 4.14 we show such a single stochastic trajectory for the $\mathbf{A} + \mathbf{B} \to \mathbf{C} + \mathbf{D}$ reaction as an example. In section 4.2.3 we shall deal with the exact simulation of trajectories in a network of K reactions and discuss the approximations leading to continuous variables $\vec{\mathcal{X}}(t) = \mathbf{x}(t)$ as used in stochastic differential equations and Fokker-Planck equations.

Chemical master equations. The general forward master equation has been written in the form

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; \sum_m \Big(W(n|m,t)P_m(t) \;-\; W(m|n,t)P_n(t) \Big)\,, \tag{3.83}$$

where we accounted for the fact that transition probabilities may be time dependent in certain cases. Here we shall assume that the transition matrix W is time independent. Before we deal with the general single reaction case we write down the formalism for two specific example: the irreversible monomolecular conversion and the reversible bimolecular conversion reaction.

Three conditions are implicitly hidden in the master equation that are readily adapted for applications to chemical kinetics [169, pp.420,421]:
Condition 1: If $\vec{\mathcal{X}}(t) = \mathbf{n}$, then the probability that *exactly one reaction* $\mathbf{R}$ will occur in the system within the time interval $[t, t + \mathrm{dt}[$ is equal to

$$\chi \; h(\mathbf{n}) \; \mathrm{dt} \; + \; o(\,\mathrm{dt}) \; ,$$

where $o(\,\mathrm{dt})$ denotes terms that approach zero with dt faster than dt.
Condition 2: If $\vec{\mathcal{X}}(t) = \mathbf{n}$, then the probability that *no* reaction will occur within the time interval $[t, t + \mathrm{dt}[$ is equal to

$$1 \; - \; \chi \; h(\mathbf{n}) \; \mathrm{dt} \; + \; o(\,\mathrm{dt}) \; .$$

Condition 3: The probability of more than one reaction occurring in the system within the time interval $[t, t + \mathrm{dt}[$ is of order $o(\,\mathrm{dt})$.

Based on the three conditions an analytical description can be derived for the evolution of the population vector $\vec{\mathcal{X}}(t)$. The initial state of the system at some initial time t_0 is fixed: $\vec{\mathcal{X}}(t_0) = \mathbf{n}_0$. We express the probability $P(\mathbf{n}, t + \mathrm{dt}|\mathbf{n}_0, t_0)$ as the sum of the probabilities of several mutually exclusive and collectively exhaustive *routes* from $\vec{\mathcal{X}}(t_0) = \mathbf{n}_0$ to $\vec{\mathcal{X}}(t + \mathrm{dt}) = \mathbf{n}$. These routes are distinguished from one another with respect to the event that happened in the last time interval $[t, t + \mathrm{dt}[$:

$$\begin{aligned} P(\mathbf{n}, t + \mathrm{dt}|\mathbf{n}_0, t_0) \; = \; & P(\mathbf{n}, t|\mathbf{n}_0, t_0) \; \times \; (1 - \gamma \, h(\mathbf{n}) \, \mathrm{dt} \; + \; o(\,\mathrm{dt})) \; + \\ & + \; P(\mathbf{n} - \mathbf{s}, t|\mathbf{n}_0, t_0) \; \times \; \Big(\gamma \, h(\mathbf{n} - \mathbf{s}) \, \mathrm{dt} \; + \; o(\,\mathrm{dt})\Big) \; + \\ & + \; o(\,\mathrm{dt}) \; . \end{aligned} \tag{4.70}$$

In few cases it is possible to derive an exact solution for the time evolution of the probability function $P(\mathbf{n}, t|\mathbf{n}_0, t_0)$, for example through solution of the chemical master equation (section 4.3), but a deterministic function for the differential change of the probability for $t \geq t_0$ is readily obtained. The three different routes from $\vec{\mathcal{X}}(t_0) = \mathbf{n}_0$ to $\vec{\mathcal{X}}(t + \mathrm{dt}) = \mathbf{n}$ are obvious from the balance equation (4.70):
(i) One route from $\vec{\mathcal{X}}(t_0) = \mathbf{n}_0$ to $\vec{\mathcal{X}}(t + \mathrm{dt}) = \mathbf{n}$ is given by the first term on the right-hand side of the equation: *No reaction* is occurring in the time interval $[t, t + \mathrm{dt}[$ and hence $\vec{\mathcal{X}}(t) = \mathbf{n}$ was fulfilled at time t. The joint probability for route (i) is therefore *the probability to be in* $\vec{\mathcal{X}}(t) = \mathbf{n}$ *conditioned by* $\vec{\mathcal{X}}(t_0) = \mathbf{n}_0$ times *the probability that no reaction has occurred in* $[t, t + \mathrm{dt}[$. In other words, the probability for this route is the probability to go from $\mathbf{n}_0$ at time t_0 to $\mathbf{n}$ at time t and to stay in this state during the next interval dt.
(ii) An alternative route from $\vec{\mathcal{X}}(t_0) = \mathbf{n}_0$ to $\vec{\mathcal{X}}(t + \mathrm{dt}) = \mathbf{n}$ is accounted for by the second term on the right-hand side of the equation: *Exactly one reaction* $\mathbf{R}$ is occurring in the time interval $[t, t + \mathrm{dt}[$ and hence $\vec{\mathcal{X}}(t) = \mathbf{n} - \mathbf{s}$ is fulfilled

at time t. The joint probability for route (ii) is therefore *the probability to be in* $\vec{\mathcal{X}}(t) = \mathbf{n} - \mathbf{s}$ *conditioned by* $\vec{\mathcal{X}}(t_0) = \mathbf{n}_0$ times *the probability that exactly one reaction* $\mathbf{R}$ *has occurred in* $[t, t + \mathrm{dt}[$. In other words, the probability for this route is the probability to go from $\mathbf{n}_0$ at time t_0 to $\mathbf{n} - \mathbf{s}$ at time t by undergoing a reaction yielding $\mathbf{n}$ during the next interval dt.
(iii) The third possibility – neither *no* reaction nor *exactly one* reaction – must inevitably invoke *more than one reaction within the time interval* $[t, t + \mathrm{dt}[$. The probability for such events, however, is $o(\mathrm{dt})$ or of measure zero.
All routes (i) and (ii) are mutually exclusive since different events are taking place within the last interval $[t, t + \mathrm{dt}[$.

The last step to derive the *chemical master equation* is straightforward: $P(\mathbf{n}, t|\mathbf{n}_0, t_0)$ is subtracted from both sides in equation (4.70), then both sides are divided by dt, the limit $\mathrm{dt} \downarrow 0$ is taken, all $o(\mathrm{dt})$ terms vanish and finally we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} P(\mathbf{n}, t|\mathbf{n}_0, t_0) \;=\; \gamma\, h(\mathbf{n}-\mathbf{s})\, P(\mathbf{n}-\mathbf{s}|\mathbf{n}_0, t_0) - \gamma\, h(\mathbf{n})\, P(\mathbf{n}, t|\mathbf{n}_0, t_0)\,. \quad (4.71)$$

Initial conditions are required to calculate the time evolution of the probability $P(\mathbf{n}, t|\mathbf{n}_0, t_0)$ and for sharp initial conditions we can easily express them in the form

$$P(\mathbf{n}, t_0|\mathbf{n}_0, t_0) \;=\; \delta_{\mathbf{n},\mathbf{n}_0} \;=\; \begin{cases} 1\,, & \text{if } \mathbf{n} = \mathbf{n}_0\,, \\ 0\,, & \text{if } \mathbf{n} \neq \mathbf{n}_0\,, \end{cases} \quad (4.71')$$

which is precisely the initial condition used in the derivation of equation (4.70): $P\big(n_k, t_0|n_k^{(0)}, t_0\big) = \delta(n_k - n_k^{(0)})\,\forall\; k$. The assumption of extended initial distributions is, of course, also possible but the corresponding master equations become more sophisticated.

Two examples. First we write down the master equation for the simple monomolecular chemical reaction (4.1c),

$$\mathsf{A} \quad \xrightarrow{\;k\;} \quad \mathsf{B}\,,$$

with the constraint of constant total number of molecules. The two random variables $\mathcal{X}_\mathsf{A}$ and $\mathcal{X}_\mathsf{B}$ fulfil the condition

$$\mathcal{X}_\mathsf{A}(t) \;+\; \mathcal{X}_\mathsf{B}(t) \;=\; n_0\;.$$

For $\mathcal{X}_\mathsf{A} = n$ and $\mathcal{X}_\mathsf{B} = n_0 - n$ it is straightforward to write down the single step master equation for the initial conditions $P(n, 0) = \delta(n - n_0)$

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; k\,(n+1)\,P_{n+1}(t) \;-\; k\,n\,P_n(t)\,,\; n \;=\; 0, 1, \ldots, n_0\;,$$

with $P_{n+1}(t) = 0 \; \forall \; t \in \mathbb{R}_{\geq 0}$, which is identical with the simple death master equation. Further discussion of the equation and its solution is found in section 4.3.2.1. For direct comparison with birth-and-death processes it is interesting to characterize the chemical elementary steps in this simple case as step-up and step-down transitions: $w_n^- = k\,n$ and $w_n^+ = 0$. We notice that the process has a built-in absorbing barrier at $n = 0$ since $w_0^- = 0$ and $w_0^+ = 0$ without any further assumption.

The second example deals with the stochastic description of the reversible bimolecular conversion reaction (4.34):

$$\begin{aligned} \mathbf{A} + \mathbf{B} &\xrightarrow{\;k\;} \mathbf{C} + \mathbf{D} \quad \text{and} \\ \mathbf{C} + \mathbf{D} &\xrightarrow{\;l\;} \mathbf{A} + \mathbf{B}\,. \end{aligned}$$

The four random variables, $\mathcal{X}_{\mathbf{A}}(t)$, $\mathcal{X}_{\mathbf{B}}(t)$, $\mathcal{X}_{\mathbf{C}}(t)$, and $\mathcal{X}_{\mathbf{D}}(t)$, are combined with three conservation relations

$$\begin{aligned} \mathcal{X}_{\mathbf{A}}(t) + \mathcal{X}_{\mathbf{B}}(t) + \mathcal{X}_{\mathbf{C}}(t) + \mathcal{X}_{\mathbf{D}}(t) &= \mathcal{X}_{\mathbf{A}}(0) + \mathcal{X}_{\mathbf{B}}(0) + \mathcal{X}_{\mathbf{C}}(0) + \mathcal{X}_{\mathbf{D}}(0) \\ \mathcal{X}_{\mathbf{A}}(t) - \mathcal{X}_{\mathbf{B}}(t) &= \mathcal{X}_{\mathbf{A}}(0) - \mathcal{X}_{\mathbf{B}}(0) \\ \mathcal{X}_{\mathbf{C}}(t) - \mathcal{X}_{\mathbf{D}}(t) &= \mathcal{X}_{\mathbf{C}}(0) - \mathcal{X}_{\mathbf{D}}(0) \end{aligned}$$

and leave only one degree of freedom. Again and we choose $\mathcal{X}_{\mathbf{A}}(t)$ as the independent variable: $P_n(t) = P(\mathcal{X}_{\mathbf{A}} = n)$. In order to simplify we assume as initial conditions that only **A** and **B** are present at time $t = 0$, and that they have sharp values: n_0 molecules **A**, $P_n(0) = \delta_{n,n_0}$, and b_0 molecules **B**, $P\big(\mathcal{X}_{\mathbf{B}}(0) = b\big) = \delta_{b,b_0}$, and we have $\mathcal{X}_{\mathbf{B}}(t) = \vartheta_0 + \mathcal{X}_{\mathbf{A}}(t)$ with $\vartheta_0 = b_0 - n_0$, and $\mathcal{X}_{\mathbf{C}}(t) = \mathcal{X}_{\mathbf{D}}(t) = n_0 - \mathcal{X}_{\mathbf{A}}(t)$. Under these conditions the master equation becomes

$$\begin{aligned} \frac{\mathrm{d}P_n(t)}{\mathrm{dt}} &= k\,(n+1)(\vartheta_0 + n + 1)\,P_{n+1}(t) + l\,(n_0 - n + 1)^2\,P_{n-1} - \\ &\quad - \Big(k\,n\,(\vartheta_0 + n) + l\,(n_0 - n)^2\Big)\,P_n(t)\;. \end{aligned}$$

This master equation is based on the step-up and step-down transitions

$$w_n^- = k\,n\,(\vartheta_0 + n) \quad \text{and} \quad w_n^+ = l\,(n_0 - n)^2\;,$$

which fulfil $\big(w_{n_0}^- = k\,n_0(\vartheta_0 + n_0) > 0, w_{n_0}^+ = 0\big)$ and $\big(w_0^- = 0, w_0^+ = l\,n_0^2 > 0\big)$ and sustain two in-built reflecting barriers at $n = n_0$ and $n = 0$, respectively, thus leaving $0 \leq n \leq n_0, n \in \mathbb{N}$ as the physically meaningful accessible domain. The master equation for the irreversible reaction (4.34a) has been solved and will be discussed in section 4.3.3.1. The full reversible reaction is rather very hard to solve and we dispense from further analysis because size expansion, numerical simulation, and chemical Langevin equation are to be preferred for practical purposes.

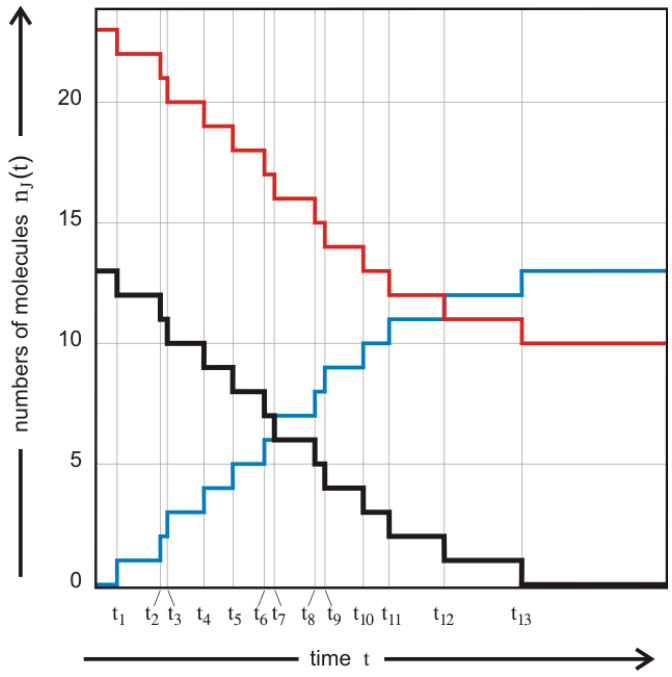


Fig. 4.14 A stochastic trajectory of the reaction $\mathbf{A}+\mathbf{B}\to\mathbf{C}+\mathbf{D}$. The figure shows a trajectory of the irreversible bimolecular conversion reaction. All four variables, $\mathcal{X}_J(t)$ with $J=\mathbf{A}$ (black), $J=\mathbf{B}$ (red), $J=\mathbf{C},\mathbf{D}$ (blue) depend on each other because of the three conservation relations discussed in the text and therefore jumps occur simultaneously at the times t_k with $k=1,\ldots,13$. In the example shown **B** molecules are present in excess. The process comes to an end when the last molecule **A** has been consumed by a reaction event or, in other words, the state $n_{\mathbf{A}}=0$ is an absorbing barrier. The intervals $\Delta t_k = t_k - t_{k-1}$ follow a Poisson distribution. Initial condition: $n_{\mathbf{A}}(0)=13$, $n_{\mathbf{B}}(0)=23$, and $n_{\mathbf{C}}=n_{\mathbf{D}}=0$.

Chemical master equation and stoichiometry. In order to describe reactants and reaction products we introduce a random vector $\vec{\mathcal{X}}$ with M components for the M molecular species:

$$\vec{\mathcal{X}} = (\mathcal{X}_1,\ldots,\mathcal{X}_M) \quad \text{with} \quad \mathcal{X}_i \in \mathbb{N}\ .$$

In the single step birth and death master equation (3.97) the variable n was undergoing two classes of changes only: step up $n\to(n+1)$ and step down $n\to(n-1)$. Dealing with a single chemical reaction at sufficient temporal resolution implies also exclusively single step transitions of the reaction $\mathbf{R}$, but they involve in general all reactants and product species: $\mathbf{n}\to\mathbf{n}\pm\mathbf{s}$. In

other words, if a reaction $\mathbf{R}$ occurs at time t, then the random vector $\vec{\mathcal{X}} \in \mathbb{N}^M$ changes according to stoichiometry: $\vec{\mathcal{X}}(t) = \vec{\mathcal{X}}(t - \mathrm{dt}) + \mathbf{s}$. Fundamental in stochastic reaction kinetics – as in the deterministic case – is the rate function $v(\mathbf{n}) = \kappa\, h(\mathbf{n})$,[37]
which takes on the form

$$v(\mathbf{n})^{(\mathrm{MA})} \;=\; \kappa\, h(\mathbf{n}) \;=\; \kappa \prod_{i=1}^{M} \frac{n_i!}{(n_i - \nu_i)!}\ . \tag{4.72}$$

The expression differs from the deterministic rate functions, $v^{(\mathrm{MA})}$ in equation (4.6), in two aspects: (i) the stochastic rate coefficient is denoted by κ but as outlined in section 4.1.4 stochastic and deterministic rate parameters are almost always assumed to be the same, $\kappa = \gamma = k$,[38] and (ii) the functions $h(\mathbf{n})$ are different, because stoichiometric coefficients $\nu_i > 1$ require explicit consideration of individual molecules at low particle numbers: The rate is proportional to the number of distinct subsets of molecules that can perform the reaction.[39]

Alternatively, we may consider the reaction as a sequence of events modeled by a continuous time Markov chain were $p\,(\tau|\mathbf{n}, t)\mathrm{d}\tau$ is the so-called *next-reaction density function*, which expresses the probability that the next reaction $\mathbf{R}$ will occur in the infinitesimal time interval $[t + \tau, t + \tau\mathrm{d}\tau[$. Every reaction changes the particle numbers by $\mathbf{s}$ and individual trajectories or sample paths follow an equation with changes at random time instances:

$$\vec{\mathcal{X}}(t) \;=\; \vec{\mathcal{X}}(0) \;+\; \mathbf{s}\; \mathcal{Y}\!\left(\int_0^t \mathrm{d}\tau\, \gamma_{\mathbf{n}(\tau)} \right)\ , \tag{4.73}$$

where $\mathcal{Y}(t)$ is an unit-rate Poisson process with the probability density $\pi_k(\tau) = e^{-\tau}\tau^k/k!$. The unit Poisson process $\mathcal{Y}(t)$ is a counting process and provides the times when the jumps in the variable $\mathcal{X}(t)$ occur whereas the stoichiometric parameters $\mathbf{s} = \boldsymbol{\nu}' - \boldsymbol{\nu}$ give the size including the sign. In this

---

[37] The rate function $v(\mathbf{n}) = \kappa\, h(\mathbf{n})$ is also known as *intensity function* or *propensity function*. The product term in equation (4.72) results from combinatorics of molecular encounters leading to a binomial coefficient $\binom{n_i}{\nu_i}$. Here, the factor $1/(\nu_i)!$ has been absorbed in the stochastic rate coefficient κ in order to obtain an expression that parallels deterministic kinetics as much as possible. We remark that the alternative notation with intact binomial factors and $\kappa \Leftrightarrow \kappa \cdot \nu_i!$ is also common.

[38] Unless stated otherwise we shall use γ or (k_j, l_j) also as rate parameters in the stochastic models. In order to take care of concentration units commonly [mol·l^{-1}], the deterministic rate parameters have to be converted to dimensionless units counting particle numbers: $k \to k/(V\, N_\mathrm{L})^{|\nu|-1}$ (see also section 4.6).

[39] The dimerization $2\mathbf{X} \to \mathbf{X}_2$, for example, yields $v^{(\mathrm{MA})} = k\, x^2$ in the deterministic case and $v_n^{(\mathrm{MA})} = \gamma\, n(n-1)$ in the stochastic case. The two expressions become identical in the limit of large numbers $\lim x \to \infty = \lim n \to \infty$ where we have $x = n$.

form both the master equation and the equation for the trajectories can be extended to mechanisms with multiple reaction steps (section 4.2.2).

Master equation and deterministic kinetics. The chemical master equation (4.71) allows for computation of expectation values. We illustrate first by means of our simple example $\mathbf{A} \to \mathbf{B}$:

$$\begin{aligned}
\sum_{n=0}^{n_0} n \frac{\mathrm{d}P_n(t)}{\mathrm{dt}} &= \frac{\mathrm{d}}{\mathrm{d}t} \sum_{n=0}^{n_0} n\, P_n(t) = \frac{\mathrm{d}\langle n \rangle}{\mathrm{dt}} = \\
&= \sum_{n=0}^{n_0} n \Big(k\,(n+1) P_{n+1}(t) - k\, n P_n(t) \Big) = \\
&= k \sum_{n=0}^{n_0} n\,(n+1) P_{n+1}(t) - k \sum_{n=0}^{n_0} n^2\, P_n(t) = \\
&= k \sum_{n'=n+1=1}^{n_0+1} (n'-1)\, n' P_{n'}(t) - k \sum_{n=0}^{n_0} n^2\, P_n(t) = \\
&= k \left\langle n^2 \right\rangle - k \langle n \rangle - k \left\langle n^2 \right\rangle = -k \langle n \rangle \ .
\end{aligned}$$

In this case the macroscopic rate equation is readily derived from the master equation through interpretation of the expectation value: $\langle n \rangle$ is a real number and up to the factor $1/(V \cdot N_\mathrm{L})$ represents the concentration of a molecular species: $x = \langle n \rangle / (V \cdot N_\mathrm{L}) \in \mathbb{R}_{\geq 0}$, and $\mathrm{d}x/\,\mathrm{dt} = -k\,x$.

Coincidence of the expectation value $\langle n \rangle$ and the deterministic particle number $\widehat{n} = x\,(V \cdot N_\mathrm{L})$ is restricted to cases where the reaction rate function is linear. A comparison of linear and nonlinear cases has already been presented and discussed in the closely related situation occurring with jump moments (sections 3.2.3.1 and 3.2.3.2).

In nonlinear examples the same procedure yields the deterministic equation form the master equation through multiplication of both sides by n_i and summation over all values of $\mathbf{n}$ in the limit of large particle numbers:[40]

$$\frac{\mathrm{d}\langle n_i(t) \rangle}{\mathrm{dt}} = s_i \cdot \langle \gamma_\mathbf{n} \rangle \quad \Longrightarrow \quad \frac{\mathrm{d}x_i(t)}{\mathrm{dt}} = s_i\, v\big(\mathbf{x}(t)\big) \ . \tag{4.74}$$

Accordingly, the master equation takes on the form of the conventional kinetic equations in this limit. It is important, however, to stress that it is not sufficient that the total number of particles is large, every molecular species $\mathbf{X}_i$ has to be present in sufficient amounts at all times t in order to make the influence of fluctuations sufficiently small.

[40] Because of the in-built or *natural* barriers its makes no difference whether the summation is running over a finite or infinite state space.

4.2.2 Stochastic chemical reaction networks

Although most studies on stochastic chemical reaction networks (SCRNs) are done by means of computer simulation, the combined analytical and numerical approach is more promising since it can give answers to general questions that cannot be addressed by pure computer work. As an example we mention the generalization of the deficiency zero theorem to master equations [9]. General texts on stochastic modeling are found in [462, 463]. We begin here with the multivariate chemical master equation and refer to section 4.6 for numerical simulations.

Multivariate chemical master equation. The master equation for many variables is readily derived by an extension of equation (4.71) to M molecular species involved in K reactions:

$$\frac{\mathrm{d}P_{\mathbf{n}}(t)}{\mathrm{dt}} \;=\; \sum_{\boldsymbol{\mu}=1}^{K} \chi_{\boldsymbol{\mu}}(\mathbf{n}-\mathbf{s}_{\boldsymbol{\mu}})\, P_{\mathbf{n}-\mathbf{s}_{\boldsymbol{\mu}}}(t) \;-\; P_{\mathbf{n}}(t) \sum_{\boldsymbol{\mu}=1}^{K} \chi_{\boldsymbol{\mu}}(\mathbf{n}) \;, \tag{4.75}$$

where we introduced vector notation: $\mathbf{n} = (n_1, \ldots, n_M)^{\mathrm{t}}$ for the particle numbers and $\mathbf{s}_{\boldsymbol{\mu}} = \boldsymbol{\nu}'_{\boldsymbol{\mu}} - \boldsymbol{\nu}_{\boldsymbol{\mu}}$ with $\boldsymbol{\nu}'_{\boldsymbol{\mu}} = (\nu'_{1\boldsymbol{\mu}}, \ldots, \nu'_{M\boldsymbol{\mu}})^{\mathrm{t}}$, and $\boldsymbol{\nu}_{\boldsymbol{\mu}} = (\nu_{1\boldsymbol{\mu}}, \ldots, \nu_{M\boldsymbol{\mu}})^{\mathrm{t}}$ for the stoichiometric coefficients. The index '$\boldsymbol{\mu}$' refers to individual reactions, $\chi_{\boldsymbol{\mu}}(\mathbf{n}) = \gamma_{\boldsymbol{\mu}} \cdot h_{\boldsymbol{\mu}}(\mathbf{n})$ is the rate function, and $\nu_{i\boldsymbol{\mu}}$ and $\nu'_{i\boldsymbol{\mu}}$ are the stoichiometry coefficients for species $\mathbf{X}_i$ in the j-th reaction. For general considerations it is simpler to avoid combinations of reactions and, for example, to treat a reversible reaction a two reaction steps. Stochastic mass action kinetics for the reaction $\mathbf{R}_{\boldsymbol{\mu}}$ is modeled by the rate function

$$\chi_{\boldsymbol{\mu}}(\mathbf{n}) \;=\; k_{\boldsymbol{\mu}} \prod_{i=1}^{M} \frac{n_i!}{(n_i - \nu_{i\boldsymbol{\mu}})!} \;=\; k_{\boldsymbol{\mu}}\, \frac{\mathbf{n}!}{(\mathbf{n}-\boldsymbol{\nu}_{\boldsymbol{\mu}})!}\,;\; \boldsymbol{\mu} = 1, \ldots, K \;, \tag{4.72'}$$

where we applied multi-index notation in the last expression. Obviously we identify $k_{\boldsymbol{\mu}} \Leftrightarrow \gamma_{\boldsymbol{\mu}}$ and $h_{\boldsymbol{\mu}}(\mathbf{n}) \Leftrightarrow \mathbf{n}!/(\mathbf{n}-\boldsymbol{\nu}_{\boldsymbol{\mu}})!$.

A stationary distribution if it exists has to fulfil

$$\overline{P}_{\mathbf{n}} \sum_{\boldsymbol{\mu}=1}^{K} \chi_{\boldsymbol{\mu}}(\mathbf{n}) \;=\; \sum_{\boldsymbol{\mu}=1}^{K} \chi_{\boldsymbol{\mu}}(\mathbf{n}-\mathbf{s}_{\boldsymbol{\mu}})\, \overline{P}_{\mathbf{n}-\mathbf{s}_{\boldsymbol{\mu}}} \quad \forall \;\; \mathbf{n} \in \Omega \;. \tag{4.76}$$

The system of equations (4.76) can be solved as shown in section 3.2.3.2.

Now we can also generalize the expression for the trajectory to the reaction network

$$\vec{\mathcal{X}}(t) \;=\; \vec{\mathcal{X}}(0) + \;\sum_{\boldsymbol{\mu}=1}^{K} s_{\boldsymbol{\mu}}\, \mathcal{Y}_{\boldsymbol{\mu}}\Big(\int_0^t \mathrm{d}\tau\, \gamma_{\boldsymbol{\mu}}\Big(\vec{\mathcal{X}}(\tau)\Big)\Big) \;, \tag{4.73'}$$

where the processes $\mathcal{Y}_{\boldsymbol{\mu}}(t)$ are independent unit-rate Poisson processes. Examples of reaction networks will be discussed in section 4.6.

Stochastic version of the deficiency zero theorem. Finally, we mention that the deficiency zero theorem has been extended to stochastic chemical reaction networks [9]: Assume a stochastic reaction network $\{\mathcal{S}, \mathcal{C}, \mathcal{R}\}$ with rate functions (4.72'),

$$\chi_{\boldsymbol{\mu}}(\mathbf{n}) \;=\; k_{\boldsymbol{\mu}} \prod_{i=1}^{M} n_i(n_i-1)\dots(n_i-\nu_{i\boldsymbol{\mu}}+1) \;,$$

for which the associated deterministic mass-action system with the same rate functions, $k_{\boldsymbol{\mu}}$; $\boldsymbol{\mu} = 1, \dots, K$, has a complex-balanced equilibrium $\overline{\mathbf{x}} \in \mathbb{R}^M_{>0}$. Then the stochastically modeled network sustains a stationary probability distribution, which is a product of Poisson distributions provided the variables x_i or n_i are independent:

$$\overline{P}_{\mathbf{n}} \;=\; \prod_{i=1}^{M} \frac{\bar{x}_i^{n_i}}{n_i!}\, e^{-\bar{x}_i} \;,\; \mathbf{n} \in \mathbb{N}^M \;, \tag{4.77}$$

If the domain of the variable $\mathbf{n}$ in $\mathbb{N}^M$ is irreducible, then (4.77) is the unique stationary distribution.

For dependent variables, for example $a = x$ and $b = n_0 - x$ in the reversible reaction $\mathbf{A} \rightleftharpoons \mathbf{B}$ with k and l as reaction rate parameters, the linear dependencies have to be eliminated, for example by setting $n_1 = n, n_2 = n_0 - n$, and with $\bar{x}_1 = \bar{x}$ and $k\bar{x} = l(n_0 - \bar{x})$ we obtain

$$\overline{P}_{\mathbf{n}} \;=\; N\,\frac{\bar{x}^n}{n!}\,\frac{(n_0-\bar{x})^{n_0-n}}{(n_0-n)!} \;=\; N\,\frac{n_0^{n_0}}{n_0!}\binom{n_0}{n}\frac{k^{n_0-n}\,l^n}{(k+l)^{n_0}} \;,\; N \;=\; \frac{n_0!}{n_0^{n_0}} \;,$$

leading to

$$\overline{P}_{\mathbf{n}} \;=\; \binom{n_0}{n}\frac{1}{(k+l)^{n_0}}\,k^{n_0-n}\,l^n \;. \tag{4.78}$$

For this reaction the distribution is binomial. Another example, the addition or association reaction $\mathbf{A} + \mathbf{B} \rightleftharpoons \mathbf{C}$, is discussed in section 4.3.3.2.

In case the domain is not irreducible the situation is more involved (figure 4.15). Then, there exist two or more closed, irreducible communicating equivalence classes, which have their own probability densities:

$$\overline{P}^{(\Gamma)}_{\mathbf{n}} \;=\; N_\Gamma \prod_{i=1}^{M} \frac{\bar{x}_i^{n_i}}{n_i!} \;,\; n_i \,\in\, \Gamma \tag{4.79}$$

with N_Γ being a normalization factor. Such a situation occurs in reactions involving two or more molecules of the same species, for example in the bimolecular conversion reaction $2\mathbf{A} \rightleftharpoons 2\mathbf{B}$ (figure 4.15). For odd total numbers of molecules, $n_0 = 2\mu+1$ with $\mu \in \mathbb{N}$, there are two systems of states $(n_{\mathbf{A}}, n_{\mathbf{B}})$

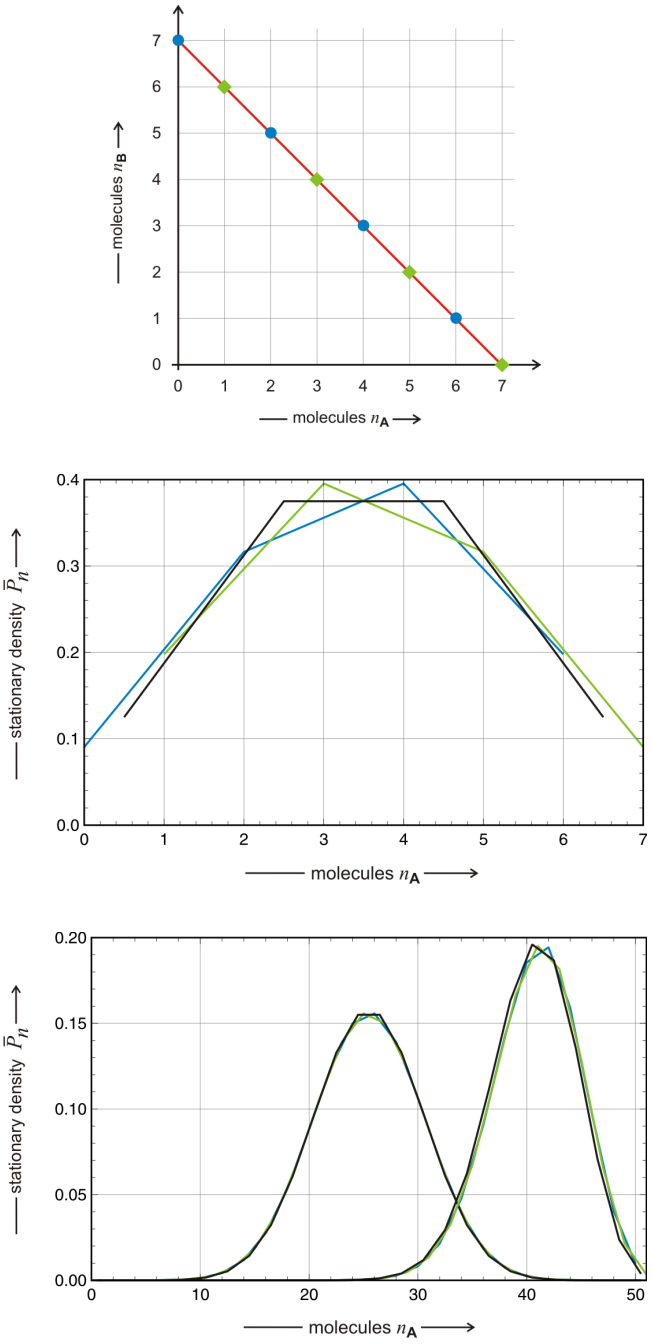


Fig. 4.15 Continued on next page.

Fig. 4.15 Irreducible communicating equivalence classes. The upper part of the figure sketches the state space for the reaction $2\mathbf{A} \underset{l}{\overset{k}{\rightleftharpoons}} 2\mathbf{B}$ with $n_{\mathbf{A}}(0) + n_{\mathbf{B}}(0) = 7$. Because of the simultaneous conversion of two molecules the domain is partitioned into two closed, irreducible communicating classes: $\Gamma_1 = \{(0,7),(2,5),(4,3),(6,1)\}$ (blue circles) and $\Gamma_2 = \{(1,6),(3,4),(5,2),(7,0)\}$ (green diamonds). The picture in the middle compares the probability densities of both irreducible classes (Γ_1, blue and Γ_2, green; $k = 2, h = 2$) with the density of the corresponding case with an even number of molecules ($n_{\mathbf{A}} + n_{\mathbf{B}} = 6$, black) that has only one irreducible class (In order to be able to compare directly the last curve has been shifted along the abscissa axis by $\Delta n = 1/2$). The diagram at the bottom compares the probability densities for $n_{\mathbf{A}}(0) + n_{\mathbf{B}}(0) = 51$ and 50, respectively, for two parameter choices: $k = 2, l = 2$ and $k = 1, l = 4$.

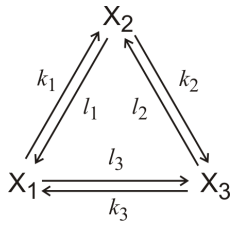


Fig. 4.16 The monomolecular triangle reaction. The sketch shows the fully reversible mechanism. For weak reversibility one cycle, either $(k_1, k_2, k_3) > 0$ or $(l_1, l_2, l_3) > 0$, is sufficient.

or equivalence classes that do not communicate:

$$\begin{aligned}(0, n_0) &\rightleftharpoons (2, n_0 - 2) \rightleftharpoons (4, n_0 - 4) \rightleftharpoons \ldots \rightleftharpoons (n_0 - 1, 1) \\ (1, n_0 - 1) &\rightleftharpoons (3, n_0 - 3) \rightleftharpoons (5, n_0 - 5) \rightleftharpoons \ldots \rightleftharpoons \quad (n_0, 0)\ .\end{aligned}$$

Although the distinction of irreducible classes is highly important for methodological reasons, it is hardly of empirical significance for chemical reactions, because the probability densities for different classes are almost identical already at fairly small particle numbers like $n_0 = 51$ in figure 4.15. This does not preclude importance in biology where the numbers of regulatory molecules can be extremely small.

The monomolecular triangle reaction. The monomolecular triangle reaction (figure 4.16) is on of the two simplest mechanisms with two independent degrees of freedom. It is straightforward to check that the deficiency of the mechanism is $\delta = 0$ and accordingly equation (4.77) applies and the stationary probability distributions can be calculated. The kinetic equations are

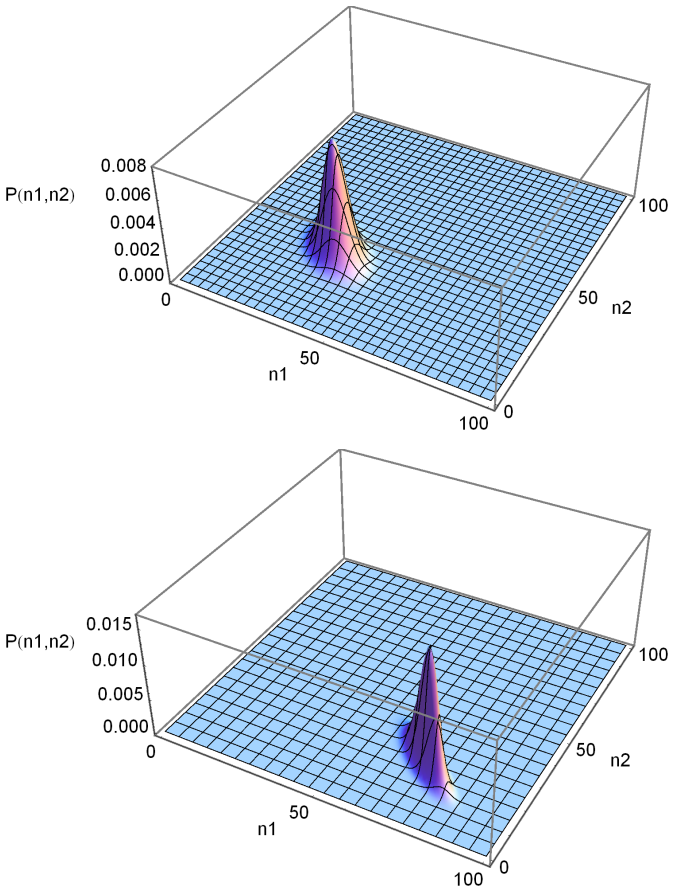


Fig. 4.17 Stationary density of the monomolecular triangle reaction. The plots show the stationary joint density $\overline{P}_{n_1,n_2}$ of the triangle reaction, $\mathbf{X}_1 \rightleftharpoons \mathbf{X}_3 \rightleftharpoons \mathbf{X}_3 \rightleftharpoons \mathbf{X}_1$. The upper plot presents the density for the symmetric case, $k_1 = k_2 = k_3 = l_1 = l_2 = l_3 = 1$, and the lower plot shows an asymmetric example: $k_1 = 1.0$, $k_2 = 2.0$, $k_3 = 10.0$, $l_1 = 1.0$, $l_2 = 0.2$, and $l_3 = 0.1$.

$$\begin{aligned}
\frac{\mathrm{d}x_1}{\mathrm{dt}} &= -(k_1 + l_3)\, x_1 + l_1\, x_2 + k_3\, x_3 \ , \\
\frac{\mathrm{d}x_2}{\mathrm{dt}} &= -(k_2 + l_1)\, x_2 + l_2\, x_3 + k_1\, x_1 \ , \quad \text{and} \\
\frac{\mathrm{d}x_3}{\mathrm{dt}} &= -(k_3 + l_2)\, x_3 + l_3\, x_1 + k_2\, x_2 \ .
\end{aligned} \tag{4.80}$$

The sum of the concentrations, $c(t) = x_1(t) + x_2(t) + x_3(t)$, fulfils a conservation relation: $c(t) = c_0 = \text{const}$, and the stationary concentrations defined

by $\mathrm{d}x_1/\,\mathrm{dt} = \mathrm{d}x_2/\,\mathrm{dt} = \mathrm{d}x_3/\,\mathrm{dt} = 0$ are readily calculated:

$$\begin{aligned} \bar{x}_1 &= (k_2k_3 + k_3l_1 + l_1l_2)\,\frac{c_0}{\Sigma}\,, \\ \bar{x}_2 &= (k_3k_1 + k_1l_2 + l_2l_3)\,\frac{c_0}{\Sigma}\,, \\ \bar{x}_3 &= (k_1k_2 + k_2l_3 + l_3l_1)\,\frac{c_0}{\Sigma} \quad \text{with} \\ \Sigma &= k_1k_2 + k_2k_3 + k_3l_1 + k_1l_2 + k_2l_3 + k_3l_1 + \\ &\quad + l_1l_2 + l_2l_3 + l_3l_1\,, \end{aligned} \tag{4.81}$$

which yields $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = c_0/3$ for the symmetric case, $k_1 = k_2 = k_3 = k$ and $l_1 = l_2 = l_3 = l$. For the thermodynamic equilibrium with the concentrations $[\mathbf{X}_i] = \bar{x}_i$ we have the condition

$$K_1 = \frac{k_1}{l_1}\,,\; K_2 = \frac{k_2}{l_2}\,,\; K_3 = \frac{k_3}{l_3}, \quad \text{and} \quad K_1\,K_2\,K_3 = \frac{k_1\,k_2\,k_3}{l_1\,l_2\,l_3} = 1\,. \tag{4.82}$$

The stationary distributions is calculated from equation (4.77):

$$\overline{P}_{n_1,n_2,n_3} = N\,\frac{\bar{n}_1^{n_1}}{n_1!}\,\frac{\bar{n}_2^{n_2}}{n_2!}\,\frac{\bar{n}_3^{n_3}}{n_3!} \quad \text{with} \quad \bar{n}_k = \bar{x}_k\,N_\mathrm{L}V\,,\; k = 1,2,3\,,$$

where the stationary concentrations $\bar{x}_k$ have been converted into stationary particle numbers $\bar{n}_k$. Figure 4.17 shows plots of the two-dimensional probability density that is centered – as expected – around the stationary point.

The cyclic closure of the mechanism introduces one constraint into the equilibrium or rate parameters. In addition, we see immediately that existence of a thermodynamic equilibrium requires that none of the six rate parameters vanishes: $(k_1, k_2, k_3, l_1, l_2, l_3) > 0$, and this is a consequence of the principle of detailed balance, which demands that the flow of each individual reaction step vanishes: $k_1\bar{x}_1 = l_2\bar{x}_2 = 0$, $k_2\bar{x}_2 = l_3\bar{x}_3 = 0$, and $k_3\bar{x}_3 = l_1\bar{x}_1 = 0$.

Stochastic model of the Michaelis-Menten mechanism. The simple Michaelis-Menten mechanism, $\mathbf{S} + \mathbf{E} \rightleftharpoons \mathbf{S}\cdot\mathbf{E} \to \mathbf{E} + \mathbf{P}$, has been studied through solving the corresponding master equation by Péter Arányi and János Tóth [13]. They obtained an exact solution under the assumption that the reaction is assumed to occur in a sufficiently small compartment that contains only a single enzyme molecule E. We remark that single molecule enzyme kinetics became accessible by experiment through modern spectroscopic techniques (section 4.4) and we shall discuss the model again in section 4.3.4. Earlier attempts to analyze the Michaelis-Menten mechanism by stochastic methods are also acknowledged [30, 228].

The extended mechanism of Michaelis-Menten type, $\mathbf{S} + \mathbf{E} \rightleftharpoons \mathbf{S}\cdot\mathbf{E} \rightleftharpoons \mathbf{E}\cdot\mathbf{P} \rightleftharpoons \mathbf{E} + \mathbf{P}$, is readily analyzed: The system has one linkage class and consists of five species, $\mathbf{S}$, $\mathbf{E}$, $\mathbf{S}\cdot\mathbf{E}$, $\mathbf{E}\cdot\mathbf{P}$, and $\mathbf{P}$, in four complexes $\mathbf{S}+\mathbf{E}$, $\mathbf{S}\cdot\mathbf{E}$, $\mathbf{E}\cdot\mathbf{P}$,

and $\mathbf{E}+\mathbf{P}$. The rank of the stoichiometric matrix is three – five concentration variables, $[\mathbf{S}] = s$, $[\mathbf{E}] = e$, $[\mathbf{S}\cdot\mathbf{E}] = c$, $[\mathbf{E}\cdot\mathbf{P}] = d$, and $[\mathbf{E}\cdot\mathbf{P}] = p$, constrained by two conservation relations for total enzyme and total substrate plus product, and accordingly we have $\delta = 4 - 1 - 3 = 0$, the deficiency zero theorem applies, and we are dealing with one unique stable stationary state. The equilibrium concentrations are given in equation (4.21) and the probability densities can be obtained from the stochastic deficiency zero theorem (4.77). The expressions, however, are too sophisticated to be used in analytical work, and numerical calculations are of limited usefulness either, because – as said already – equilibrium conditions are rarely applied in experimental studies or in biotechnology. Stochastic Michaelis-Menten kinetics will be discussed again in section 4.3.4 and later in section 4.4 where we are dealing with single molecule enzyme kinetics.

Finally, we mention that we have presented here only the simplest examples of reactions for the purpose of illustration but the *stochastic chemical reaction network approach* (SCRN) turned out to by very useful for modeling of *real* networks in systems biology and we shall encounter examples in the forthcoming sections.

4.2.3 The chemical Langevin equation

Stochastic differential equations (SDEs) were introduced in section 3.4 as an alternative to the Chapman-Kolmogorov approach to handle stochastic processes (figure 3.1) and we introduced the chemical Langevin equation and mentioned some of its most salient features already there. The Chapman-Kolmogorov approach in its various forms – master equation, Fokker-Planck equation, and others – aims at modeling the evolution of probability densities, whereas stochastic differential or Langevin equations (SDEs) concern single trajectories. Here we present a short derivation of the Langevin equation applied to chemical kinetics [171] in order to illustrate the implicit approximations of the approach.

At first, equation (4.69) is generalized to a network of K different reactions $\mathbf{R}_j$ $(j = 1, \ldots, K)$ and we define a time interval $\Delta t = [t, t+\tau]$[41] for the recording of reaction events. The random vector of particle numbers at the beginning of the time interval is $\vec{\mathcal{X}}(t) = \mathbf{n}(t)$ and at the end of the interval we have

$$\begin{aligned} \mathbf{n}(t+\tau) &= \mathbf{n}(t) + \mathrm{S} \cdot \boldsymbol{\eta}\big(\mathbf{n}(t), \tau\big) \quad \text{or} \\ \mathrm{n}_i(t+\tau) &= \mathrm{n}_i(t) + \sum_{j=1}^{K} s_{ij}\, \eta_j\big(\mathbf{n}(t), \tau\big)\,, \; i = 1, \ldots, M\,. \end{aligned} \tag{4.83}$$

with S being the stoichiometric matrix and $\boldsymbol{\eta} = (\eta_1 \ \ldots, \eta_K)^{\mathrm{t}}$ a random vector counting the numbers of $\mathbf{R}_j$ reaction events, which occurred during the interval $[t, t+\tau]$. These random numbers obviously depend on the time-dependent particle numbers and the length of the interval: $\boldsymbol{\eta} = \boldsymbol{\eta}\big(\mathbf{n}(t), \tau\big)$.

The fully fledged computation of $\mathcal{X}_i(t+\tau)$ for arbitrary $\tau > 0$ is definitely as difficult as a solution of the corresponding master equation (4.71) but (4.83) allows for the introduction of transparent approximations provided two conditions (3.184) are fulfilled that we repeat here:

(1) the interval τ has to be sufficiently short such that the rate functions do not change appreciably during $[t, t+\tau]$:
$$\gamma_j\big(\vec{\mathcal{X}}(\theta)\big) \approx \gamma_j\big(\mathbf{n}(t)\big) \; \forall \; \theta \in [t, t+\tau] \text{ and } \forall\, j = 1 \ldots, K,$$
(2) the interval τ has to be long enough that the expected number of occurrences of reactions in each reaction channel during the time interval $[t, t+\tau]$ must be larger than one:
$$\big\langle \mathcal{P}_j\big(\gamma_j\big(\mathbf{n}(t)\big), \tau\big)\big\rangle = \gamma_j\big(\mathbf{n}(t)\big)\tau \gg 1\,, \forall\, j = 1 \ldots, K,$$
where the quantities $\mathcal{P}_j(\alpha)$ are random variables with Poissonian density (section 2.3.1).

Condition (1) simplifies substantially the equation for the stochastic trajectory: Because of the assumed constancy of $\gamma_j\big(\mathbf{n}(t)\big)$ during the interval $[t, t+\tau]$

[41] The time interval is the same for all reactions, it is predetermined, and thus it is different in nature from the time interval in equation (4.69).

the random variables $\eta_j\big(\mathbf{n}(t),\tau\big)$ become statistically independent Poisson variables $\mathcal{P}_j\big(\gamma_j\big((\mathbf{n}(t)\big),\tau\big)$ and equation (4.83) can be approximated by

$$\mathrm{n}_i(t+\tau) \;=\; \mathrm{n}_i(t) \;+\; \sum_{j=1}^{K} s_{ij}\,\mathcal{P}_j\big(\gamma_j\big(\mathbf{n}(t)\big),\tau\big)\,,\; i=1,\ldots,M\;. \qquad (4.83')$$

The second condition (2) allows to approximate the discrete Poisson variable $\mathcal{P}$ by a continuous random variable with normal distribution $\mathcal{N}(t) = \mathbf{x}(t)$ (section 2.3.3) whereby the mean and variance remain the same:

$$\mathcal{X}_i(t+\tau) \;=\; x_i(t) \;+\; \sum_{j=1}^{K} s_{ij}\,\mathcal{N}_j\big(\gamma_j\big(\mathbf{x}(t)\big)\tau,\gamma_j\big(\mathbf{x}(t)\big)\tau\big)\,,\; i=1,\ldots,M\;. \qquad (4.83'')$$

At the same time this approximation changed the originally discrete random variables $\mathcal{X}_j(t) = n_j(t)$ into continuous variables $\mathcal{X}_j(t) = x_j(t)$ on the domain of the nonnegative real numbers: $x_j \in \mathbb{R}_{\geq 0}\,\forall\, j = 1,\ldots,M$. Using the linear combination theorem for normal variables, $\mathcal{X}(m,\sigma^2) = m + \sigma\mathcal{X}(0,1)$ we can rewrite this equation and find

$$\mathcal{X}_i(t+\tau) \;=\; x_i(t) \;+\; \sum_{j=1}^{K} s_{ij}\,\gamma_j\big(\mathbf{x}(t)\big)\,\tau \;+\; \sum_{j=1}^{K} s_{ij}\,\sqrt{\gamma_j\big(\mathbf{x}(t)\big)\tau}\,\mathcal{N}(0,1)\,;\; i=1,\ldots,M\,.$$

Recalling that the probability density of the Wiener process W is the standard normal distribution we obtain

$$\mathcal{X}_i(t+\,\mathrm{dt}) \;=\; \mathcal{X}_i(t) \;+\; \sum_{j=1}^{K} s_{ij}\,\gamma_j\big(\vec{\mathcal{X}}(t)\big)\;\mathrm{dt} + \sum_{j=1}^{K} s_{ij}\,\sqrt{\gamma_j\big(\vec{\mathcal{X}}(t)\big)}\,\mathcal{N}(0,1)(\,\mathrm{dt})^{1/2}\,;$$
$$i=1,\ldots,M\,,$$

which after rewriting yields the chemical Langevin equation

$$\mathrm{d}x_i \;=\; \sum_{j=1}^{K} s_{ij}\,\gamma_j\big(\vec{\mathcal{X}}(t)\big)\;\mathrm{dt} \;+\; \sum_{j=1}^{K} s_{ij}\,\sqrt{\gamma_j\big(\vec{\mathcal{X}}(t)\big)}\,\mathrm{d}W_j(t)\,;\; i=1,\ldots,M\;. \qquad (4.84)$$

Each reaction $\mathbf{R}_j$ $(j=1,\ldots,K)$ contributes to fluctuations of particle numbers $\mathcal{X}_i$ $(i=1,\ldots,M)$ as a Wiener process W_j: The K contributions are temporarily uncorrelated, statistically independent white noise terms.

Although the two approximations appear to be contradictory (figure 3.30) since for approximation (1) τ has to be *sufficiently small* but at the same it has to be *large enough* to fulfil condition (2). Considering, however, particle numbers of 10^{20} and more that are typical in conventional chemistry we realize that there is indeed enough room to have variables, whose size allows for a sufficiently accurate approximation of a Poissonian by a normal

distribution and for which changes $\Delta \mathrm{n} = \pm 1, \pm 2 \ldots$ lead to negligibly small relative variations. These approximation method also known as τ-leaping will be discussed in more detail in section 4.6.2.

Equation (4.84) corresponds to a forward Fokker-Planck equation (see section 3.4.4), which describes the evolution of the multivariate probability density of the random vector $\vec{\mathcal{X}}(t)$ [171]:

$$\begin{aligned}\frac{\mathrm{d}P(\mathbf{x},t)}{\mathrm{dt}} &= -\sum_{i=1}^{M} \frac{\partial}{\partial x_i} \left(\Big(\sum_{k=1}^{K} s_{ij} \gamma_k(\mathbf{x}) \Big) P(\mathbf{x},t) \right) + \\ &+ \frac{1}{2} \sum_{i=1}^{M} \frac{\partial^2}{\partial x_i^2} \left(\Big(\sum_{k=1}^{K} s_{ik}^2 \gamma_k(\mathbf{x}) \Big) P(\mathbf{x},t) \right) + \\ &+ \sum_{i,j=1\,;\,i<j}^{M} \frac{\partial^2}{\partial x_i \partial x_j} \left(\Big(\sum_{k=1}^{K} s_{ik} s_{jk} \gamma_k(\mathbf{x}) \Big) P(\mathbf{x},t) \right)\end{aligned} \tag{4.85}$$

The initial conditions are $P(\mathbf{x}_0, t_0) = \delta(\mathbf{x} - \mathbf{x}_0)$.

As shown in figure 3.1 the equivalence of Langevin and Fokker-Planck equations is a bridge built by rigorous mathematics between the single trajectory and the probability density approach in chemical kinetics. This equivalence is based on the usage of continuous variables, which in case of reaction kinetics is almost always well justified by the large particle numbers in chemistry. We mention that a second bridge exists between the numerical simulation of stochastic trajectories and chemical master equations that is mediated by numerical mathematics on the level of discrete variables [169, 173].

4.3 Examples of chemical reactions

In this section we shall present exact solutions of chemical master equations for study cases from three classes of chemical reactions: (i) zero-molecular reactions in form of the flow in a reactor, (ii) monomolecular reactions with one reactant, and (iii) bimolecular reactions involving two reactants. Molecularity of a reaction is commonly reflected by the chemical rate law of reaction kinetics in form of the reaction order. In particular, we distinguish fist order and second order kinetics, which is typically observed with monomolecular and bimolecular reactions, respectively. Exceptions are conditions like excess of one reactant, which leads to an observed reaction order that is smaller than the molecularity. The most frequently encounter example are *pseudo first order reactions* (see section 4.3.3.1). Because of its fundamental importance in chemistry and biology the autocatalytic elementary step (4.1g) will be discussed in a separate section 4.3.3.3.

4.3.1 The flow reactor

The flow reactor is introduced as an experimental device that allows for investigations of systems under controlled conditions away from thermodynamic equilibrium. The establishment of a stationary state or *flow equilibrium* in a flow reactor (CFSTR or CSTR: continuous flow stirred tank reactor; figure 4.18) is a suitable case study for the illustration of the search for the solution of a birth-and-death master equation. At the same time the non-reactive flow of a single compound represents the simplest conceivable process in such a reactor. The stock solution contains **A** at the concentration $[\mathbf{A}]_{\text{in}} = \widehat{a} = \bar{a}\,[\text{mol}\cdot\text{l}^{-1}]$. The inflow concentration $\widehat{a}$ is equal to the stationary concentration $\bar{a}$, because no reaction is assumed to take place in the reactor.

Flow in case of the flow reactor in figure 4.18 is understood as volume flow and expressed in terms of the volume flow rate r that is measured in the units $[\text{l}\cdot\text{sec}^{-1}]$.[42] Accordingly the inflow of compound **A** into the reactor is $\widehat{a}\cdot r\,[\text{mol}\cdot\text{sec}^{-1}]$ expressed in concentration after instantaneous mixing with the content of the reactor. The outflow of the mixture in the reactor occurs with the same flow rate r.[43] The reactor has a volume of $V\,[\text{l}]$ and thus we have a mean residence time of $\tau_{\text{v}} = V\cdot r^{-1}\,[\text{sec}]$ of a volume element dV in the reactor.

---

[42] Volume flow is to be distinguished from *mass flow* the measure of which is a mass flow rate $\widetilde{r}\,[\text{kg}\cdot\text{sec}^{-1}]$. Mass flow is a scalar quantity, and when it is measured with respect to a unit area through which the transport takes place, it is called a *flux* $\boldsymbol{\phi}$ measured in $[\text{kg}/(\text{m}^2\cdot\text{sec}^1)]$. Flux in contrast to flow is a vector perpendicular to the reference unit area [412].

[43] The assumption of equal inflow and outflow rate is required because we are dealing with a flow reactor of constant volume V (CSTR, figure 4.18).

In- and outflow of compound **A** into and from the reactor are modeled by two formal elementary steps or pseudoreactions

$$\begin{aligned} \star &\longrightarrow \mathbf{A} \\ \mathbf{A} &\longrightarrow \oslash \ . \end{aligned} \tag{4.86}$$

In chemical kinetics the differential equations are almost always formulated in terms of molecular concentrations. For the stochastic treatment the concentrations are replaced by particle numbers, $n = a \cdot V \cdot N_{\mathrm{L}}$ with $n \in \mathbb{N}$ and N_{L} being Avogadro's number.

The particle number of **A** in the reactor is a stochastic variable $\mathcal{N}(t)$ with the probability $P_n(t) = P\big(\mathcal{N}(t) = n\big)$. The time derivative of the probability density is described by means of the master equation

$$\frac{\partial P_n(t)}{\partial t} \ = \ r\left(\widehat{n}\, P_{n-1}(t) + (n+1)\, P_{n+1}(t) - (\widehat{n}+n)\, P_n(t)\right); \ n \in \mathrm{N}\ , \tag{4.87}$$

where the flow rate could be absorbed by a redefinition of the time axis. Equation (4.87) is equivalent to a birth-and-death process with the step-up and step-down transition probabilities $w_n^+ = r\,\widehat{n}$ and $w_n^- = r\,n(t)$, respectively. Thus we have a constant birth rate and a death rate which is proportional to n. Solutions of the master equation can be found in text books listing stochastic processes with known solutions, for example [176].

Here we derive the solution by means of probability generating functions as introduced in equation (2.24) (subsection 2.2.1):

$$g(s,t) \ = \ \sum_{n=0}^{\infty} P_n(t)\, s^n \ . \tag{2.24'}$$

The initial state is included in the expression: $g_{n_0}(s,t)$ implies $P_n(0) = \delta_{n,n_0}$. Partial derivatives with respect to time t and the dummy variable s are readily computed:

$$\begin{aligned} \frac{\partial g(s,t)}{\partial t} \ &= \ \sum_{n=0}^{\infty} \frac{\partial P_n(t)}{\partial t} \cdot s^n \ = \\ &= \ r \sum_{n=0}^{\infty} \Big(\widehat{n}\, P_{n-1}(t) \ + \ (n+1)\, P_{n+1}(t) \ - \ (\widehat{n}+n)\, P_n(t)\Big)\, s^n \quad \text{and} \\ \frac{\partial g(s,t)}{\partial s} \ &= \ \sum_{n=0}^{\infty} n\, P_n(t)\, s^{n-1} \ . \end{aligned}$$

Proper collection of terms and rearrangement of summations – by taking into account that $w_0^- = 0$ – yields

$$\frac{\partial g(s,t)}{\partial t} \ = \ r\,\widehat{n} \sum_{n=0}^{\infty} \Big(P_{n-1}(t) \ - \ P_n(t)\Big)\, s^n \ + \ r \sum_{n=0}^{\infty} \Big((n+1)\, P_{n+1}(t) \ - \ n\, P_n(t)\Big)\, s^n \ .$$

Evaluation of the four infinite sums

$$\sum_{n=0}^{\infty} P_{n-1}(t)\, s^n \;=\; s \sum_{n=0}^{\infty} P_{n-1}(t)\, s^{n-1} \;=\; s\, g(s,t)\,,$$
$$\sum_{n=0}^{\infty} P_n(t)\, s^n \;=\; g(s,t)\,,$$
$$\sum_{n=0}^{\infty} (n+1)\, P_{n+1}(t)\, s^n \;=\; \frac{\partial g(s,t)}{\partial s}\,, \quad \text{and}$$
$$\sum_{n=0}^{\infty} n\, P_n(t)\, s^n \;=\; s \sum_{n=0}^{\infty} n\, P_n(t)\, s^{n-1} \;=\; s\, \frac{\partial g(s,t)}{\partial s}\;,$$

and regrouping of terms yields a linear partial differential equation of first order

$$\frac{\partial g(s,t)}{\partial t} \;=\; r\Big(\widehat{n}(s-1)\, g(s,t) \;-\; (s-1)\, \frac{\partial g(s,t)}{\partial s}\Big)\;. \tag{4.88}$$

A general method to derive solutions called *method of characteristics* exists for linear first order partial differential equation [474, pp. 390-396]. The trick is to reduce the problem of solving a PDE to the task to find solutions for an ODE. We briefly illustrate this solution technique.

In order to solve the PDE we start with a substitution[44]

$$g(s,t) \;=\; \phi(s,t)\, e^{\widehat{n}\, s}\,,\; \frac{\partial g(s,t)}{\partial t} \;=\; \frac{\partial \phi(s,t)}{\partial t}\, e^{\widehat{n}\, s}\,, \text{ and}$$
$$\frac{\partial g(s,t)}{\partial s} \;=\; \big(\widehat{n}\, \phi(s,t) \;+\; \frac{\partial \phi(s,t)}{\partial s}\big)\, e^{\widehat{n}\, s}\,, \text{ yielding}$$
$$\frac{\partial \phi(s,t)}{\partial t} \;+\; r(s-1)\, \frac{\partial \phi(s,t)}{\partial s} \;=\; 0\;.$$

We define a *characteristic manifold* consisting of curves that are defined by the tangent vector $\mathbf{v}_{\mathrm{t}} = \big(1, r(s-1)\big)$ and accordingly the characteristic curves have to satisfy the ODE:

$$\frac{\mathrm{d}s}{\mathrm{d}t} \;=\; r(s-1) \;\text{ and }\; s-1 \;=\; e^{rt} \cdot C\;,$$

where C is the integration constant. The *characteristic curves* fulfil the equation $e^{-rt}(s-1) = C$ and the solutions of the PDE take on the form

$$\phi(s,t) \;=\; f(C) \;=\; f\big(e^{-rt}(s-1)\big) \;\text{ and }\; g(s,t) \;=\; N\, f\big(e^{-rt}(s-1)\big)\, e^{\widehat{n}\, s}\;,$$

where N is a normalization factor, and f is some arbitrary function. The factor N is readily calculated for the solution with $s = 1$:

$$g(1,t) \;=\; 1 \quad \Rightarrow \quad N\, f(0)\, e^{\widehat{n}} \;=\; 1\,,\; f(0) = 1\,,\; N = e^{-\widehat{n}}\;,$$

[44] The substitution is dispensable but leads to simpler expressions.

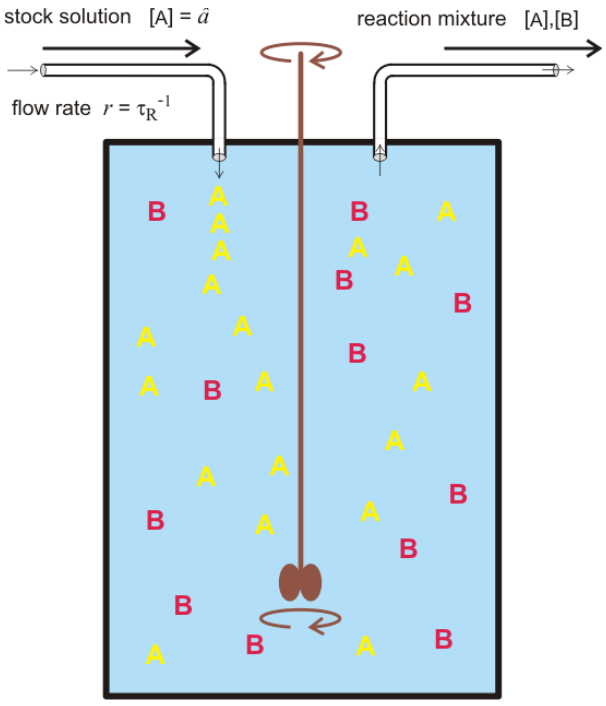


Fig. 4.18 The flow reactor. The reactor shown in the sketch is a device for experimental and theoretical chemical reaction kinetics, which is used to carry out chemical reactions in an open system. The stock solution contains materials, for example **A** at the concentration $[\mathbf{A}]_{\text{in}} = \widehat{a}$, which are usually consumed during the reaction to be studied. The reaction mixture is stirred in order to guarantee a spatially homogeneous reaction medium. Constant volume implies an outflow from the reactor that compensates in volume the inflow. The flow rate r is equivalent to the inverse mean residence time of solution in the reactor multiplied by the reactor volume, $\tau_{\text{v}}^{-1} \cdot V = r$. The reactor shown here is commonly called continuously stirred tank reactor (CSTR).

and we obtain

$$g(s,t) \;=\; f\big(e^{-rt}(s-1)\big)\, e^{\widehat{n}\,(s-1)} \;.$$

In order to determine f we use the initial condition $P_n(0) = \delta_{n,n_0}$:

$$\begin{aligned} g(s,0) \;&=\; f(s-1)\,\cdot\,\exp\big((s-1)\widehat{n}\big) \;=\; s^{n_0} \;, \\ f(\zeta) \;&=\; (\zeta+1)^{n_0}\cdot\exp\left(-\zeta\widehat{n}\right) \quad \text{with} \quad \zeta = (s-1)\,e^{-rt} \;, \\ g(s,t) \;&=\; \Big(1+(s-1)\,e^{-rt}\Big)^{n_0}\cdot\exp\big(-\widehat{n}(s-1)\,e^{-rt}\big)\cdot\exp\big(\widehat{n}(s-1)\big) \;= \\ &=\; \Big(1+(s-1)\,e^{-rt}\Big)^{n_0}\cdot\exp\big\{-\widehat{n}(s-1)\,(1-e^{-rt})\big\} \;. \end{aligned} \tag{4.89}$$

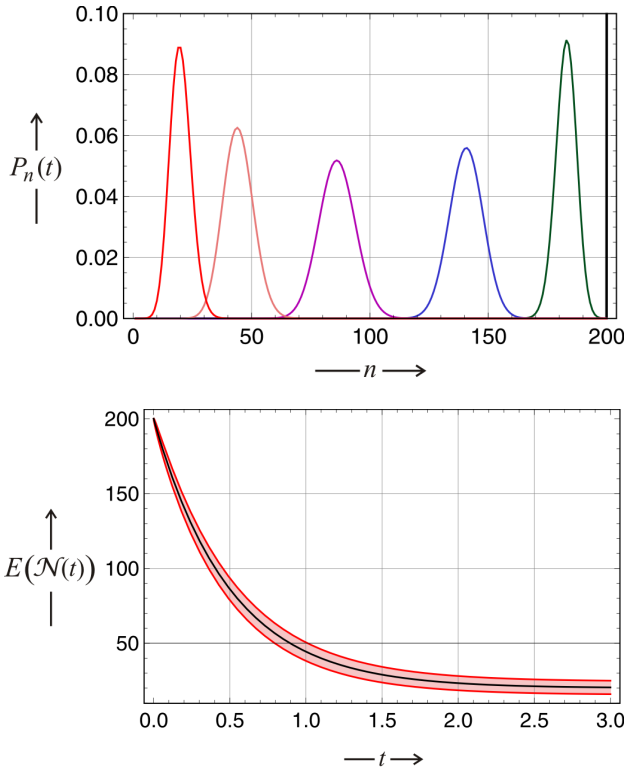


Fig. 4.19 Establishment of the flow equilibrium in the CSTR. The upper part shows the evolution of the probability density, $P_n(t)$, of the number of molecules of a compound **A** which flows through a reactor of the type illustrated in figure 4.18. The initially infinitely sharp density becomes broader with time until the variance reaches its maximum and then sharpens again until it reaches stationarity. The stationary density is a Poissonian with expectation value and variance, $E(\mathcal{N}) = \mathrm{var}(\mathcal{N}) = \widehat{n}$. In the lower part we show the expectation value $E\big(\mathcal{N}(t)\big)$ in the confidence interval $E \pm \sigma$. Parameters used: $\widehat{n} = 20$, $n_0 = 200$, and $V = 1$; sampling times (upper part): $\tau = r \cdot t = 0$ (black), 0.05 (green), 0.2 (blue), 0.5 (violet), 1 (pink), and ∞ (red).

From the generating function we compute with somewhat tedious but straightforward algebra the probability distribution

$$P_n(t) \;=\; \sum_{k=0}^{\min\{n_0,n\}} \binom{n_0}{k} \widehat{n}^{n-k} \cdot \frac{e^{-krt}\left(1-e^{-rt}\right)^{n_0+n-2k}}{(n-k)!} \cdot e^{-\widehat{n}\left(1-e^{-rt}\right)} \qquad (4.90)$$

with $n, n_0, \widehat{n} \in \mathbb{N}^0$. In the limit $t \to \infty$ we find a non vanishing contribution to the stationary probability only from the first term, $k = 0$, and obtain

$$\lim_{t\to\infty} P_n(t) \;=\; \bar{P}_n(t) \;=\; \frac{\widehat{n}^n}{n!}\,\exp\left(-\widehat{n}\right)\,. \tag{4.91}$$

This is a Poissonian distribution with parameter and expectation value $\alpha = \widehat{n}$. The Poissonian distribution has also a variance, which is numerically identical to the expectation value, $\mathrm{var}(\mathcal{N}) = E(\mathcal{N}) = \widehat{n}$, and at the stationary state the distribution of particle numbers fulfils perfectly a $\sqrt{n}$-law.

The time dependent probability distribution allows to compute the expectation value and the variance of the particle number as a function of time

$$\begin{aligned} E\big(\mathcal{N}(t)\big) \;&=\; \widehat{n} \,+\, (n_0 - \widehat{n})\cdot e^{-rt}\,, \\ \mathrm{var}\big(\mathcal{N}(t)\big) \;&=\; \big(\widehat{n} \,+\, n_0\cdot e^{-rt}\big)\cdot\big(1 \,-\, e^{-rt}\big)\,. \end{aligned} \tag{4.92}$$

As expected for linear transition probabilities the expectation value coincides with the solution curve of the deterministic differential equation

$$\frac{\mathrm{d}n}{\mathrm{dt}} \;=\; w_n^+ \,-\, w_n^- \;=\; r\,(\widehat{n} - n) \quad\text{and}\quad n(t) \;=\; \widehat{n} \,+\, (n_0 - \widehat{n})\cdot e^{-rt}\,. \tag{4.93}$$

Since we start from sharp initial densities, variance and standard deviation are zero at time $t = 0$. The qualitative time dependence of $\mathrm{var}\big(\mathcal{N}(t)\big)$, however, depends on the sign of $(n_0 - \widehat{n})$:

(i) For $n_0 \le \widehat{n}$ the standard deviation increases monotonously until it reaches the stationary value $\sqrt{\widehat{n}}$ in the limit $t \to \infty$, and

(ii) for $n_0 > \widehat{n}$ the standard deviation increases until it passes through a maximum at

$$t(\sigma_{\max}) \;=\; \frac{1}{r}\left(\ln 2 \,+\, \ln n_0 \,-\, \ln(n_0 - \widehat{n})\right)$$

and approaches the long-time value $\sqrt{\widehat{n}}$ from above.

In figure 4.19 we show an example for the evolution of the probability density (4.90). In addition, the figure contains a plot of the expectation value $E\big(\mathcal{X}(t)\big)$ inside the band $E - \sigma < E < E + \sigma$. In case of a normally distributed stochastic variable we find 68.3% of all values within this *confidence interval*. In the interval $E-2\sigma < E < E+2\sigma$ we would find even 95.4% of all stochastic trajectories (2.3.3).

4.3.2 Monomolecular chemical reactions

The reversible mono- or unimolecular chemical reaction can be split into two irreversible elementary reactions

$$\mathbf{A} \xrightarrow{k} \mathbf{B} \tag{4.94a}$$

$$\mathbf{A} \xleftarrow{l} \mathbf{B} \,, \tag{4.94b}$$

wherein k and l, are the *reaction rate parameters*, which depend on temperature, pressure, and other environmental factors. At equilibrium the rate of the forward reaction (4.94a) is precisely compensated by the rate of the reverse reaction (4.94b), $k \cdot [\mathbf{A}] = l \cdot [\mathbf{B}]$, leading to the condition for the thermodynamic equilibrium:

$$K = \frac{k}{l} = \frac{[\mathbf{B}]}{[\mathbf{A}]} \,, \tag{4.95}$$

where the parameter K is the *equilibrium constant*, which again depends on temperature, pressure, and other environmental factors like the reaction rate parameters. In an isolated or in a closed system we have a conservation law:

$$\frac{\mathcal{X}_A(t) + \mathcal{X}_B(t)}{V \cdot N_{\mathrm{L}}} = [\mathbf{A}] + [\mathbf{B}] = c(t) = c_0 = \bar{c} = \text{constant} \,, \tag{4.96}$$

where c is the constant total concentration.

The two irreversible reactions are characterized by vanishing rate parameters, $\lim l \to 0$ or $\lim k \to 0$, respectively. It is worth mentioning that zero rate parameters correspond to an instability in the Gibbs free energy at equilibrium, $\Delta G_0 = -RT \ln K$, and are incompatible with rigorous thermodynamics. Nevertheless, the assumption of irreversibility is a good approximation in cases where equilibria are lying almost completely on the side of reactants or products, respectively.

4.3.2.1 Irreversible monomolecular reaction

We start by discussing the simpler irreversible case,

$$\mathbf{A} \xrightarrow{k} \mathbf{B} \,, \tag{4.94a'}$$

which is can be modeled and analyzed in full analogy to the previous case of the flow equilibrium. We are dealing with two molecular species **A** and **B**, but the process is fully described by a single stochastic variable, $\mathcal{X}_A(t) = n$, since we have the conservation relation (4.96) $\mathcal{X}_A(t) + \mathcal{X}_B(t) = n_0$ with $n_0 = n(0)$ being the initial number of **A** molecules. The reaction is tantamount to a

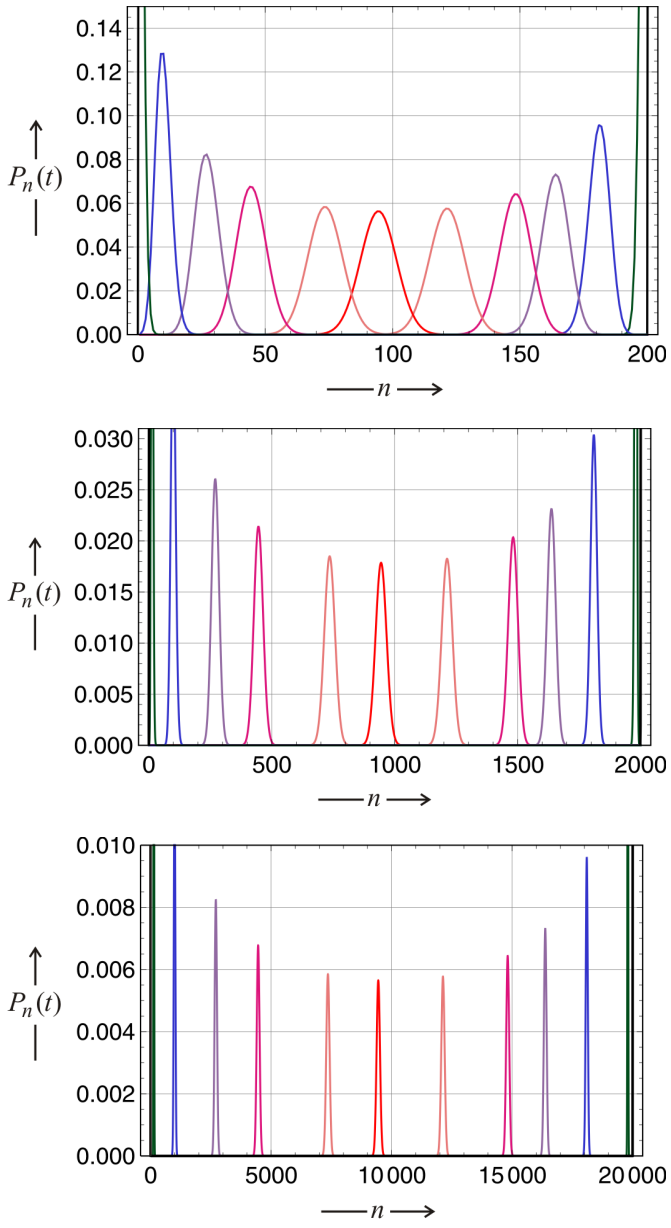


Fig. 4.20 Continued on next page.

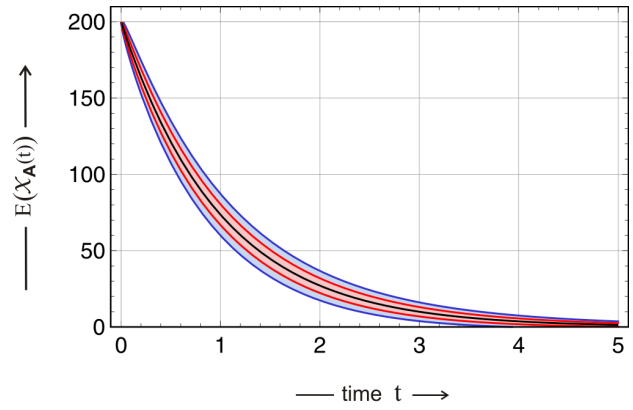


Fig. 4.20 Probability density of an irreversible monomolecular reaction. The three plots on the previous page show the evolution of the probability density, $P_n(t)$, of the number of molecules of a compound **A** which undergo a reaction **A**→**B**. The initially infinitely sharp density $P_n(0) = \delta_{n,n_0}$ becomes broader with time until the variance reaches its maximum at time $t = t_{1/2} = t_{\max} = \ln 2/k$ and then sharpens again until it approaches full transformation, $\lim_{t\to\infty} P_n(0) = \delta_{n,0}$. On this page we show the expectation value $E\big(\mathcal{X}_A(t)\big)$ and the confidence intervals $E \pm \sigma$ (68,3%,red) and $E \pm 2\sigma$ (95,4%,blue) with $\sigma^2 = \mathrm{var}\big(\mathcal{X}_A(t)\big)$ being the variance. Parameters used: $n_0 = 200$, 2000, and 20 000; $k = 1\,[t^{-1}]$; sampling times: 0 (black), 0.01 (green), 0.1 (blue), 0.2 (violet), (0.3) (magenta), 0.5 (pink), 0.75 (red), 1 (pink), 1.5 (magenta), 2 (violet), 3 (blue), and 5 (green). The initial condition for the time dependence of the expectation value was $n_0 = 200$.

single-step pure death process with $w_n^+ = 0$ and $w_n^- = k\,n$.[45] The probability density is defined by $P_n(t) = P(\mathcal{X}_A = n)$ and the time dependence obeys the master equation

$$\frac{\partial P_n(t)}{\partial t} \;=\; k\,(n+1)\,P_{n+1}(t) \;-\; k\,n\,P_n(t)\;. \tag{4.97}$$

Equation (4.97) can be solved again by means of the probability generating function,

$$g(s,t) \;=\; \sum_{n=0}^{\infty} P_n(t)\,s^n\,;\;\; |s| \le 1\;,$$

which again fulfils a linear PDE of first order

$$\frac{\partial g(s,t)}{\partial t} \;-\; k\,(1-s)\,\frac{\partial g(s,t)}{\partial s} \;=\; 0\;,$$

[45] We remark that $w_n^- = 0$ and $w_n^+ = 0$, the conditions for a natural absorbing barrier (section 5.2.1), are fulfilled at $n = 0$.

which can be solved as shown in the previous section 4.3.1 and yields

$$g(s,t) \;=\; \Big(1 \,+\, (s-1)e^{-kt}\Big)^{n_0} \;. \tag{4.98}$$

This expression is expanded in binomial form, which introduces an ordering with respect to increasing powers of s,

$$g(s,t) = (1-e^{-kt})^{n_0} + \binom{n_0}{1} s e^{-kt}(1-e^{-kt})^{n_0-1} + \binom{n_0}{2} s e^{-2kt}(1-e^{-kt})^{n_0-2} + \\ + \ldots + \binom{n_0}{n_0-1} s^{n_0-1} e^{-(n_0-1)kt}(1-e^{-kt}) + s^{n_0} e^{-n_0 kt} \;.$$

Comparison of coefficients yields the time dependent probability density

$$P_n(t) \;=\; \binom{n_0}{n} \Big(\exp\left(-kt\right)\Big)^n \Big(1-\exp\left(-kt\right)\Big)^{n_0-n} \;. \tag{4.99}$$

It is straightforward to compute the expectation value of the stochastic variable $\mathcal{X}_A$, which again coincides with the deterministic solution, and its variance

$$\begin{aligned} E\big(\mathcal{X}_A(t)\big) \;&=\; n_0\, e^{-kt} \;, \\ \mathrm{var}\big(\mathcal{X}_A(t)\big) \;&=\; n_0\, e^{-kt}\, \big(1-e^{-kt}\big) \;. \end{aligned} \tag{4.100}$$

The half-life of a population of n_0 particles,

$$t_{1/2}: \; E\{\mathcal{N}_A(t)\} \,=\, \frac{n_0}{2} \,=\, n_0 \cdot e^{-kt_{\max}} \quad \Longrightarrow \quad t_{1/2} \,=\, t_{\max} \,=\, \frac{1}{k}\ln 2 \;,$$

is the time of maximum variance or standard deviation, $\mathrm{d}\,\mathrm{var}\big(\mathcal{X}_A(t)\big)/\,\mathrm{dt} = 0$ or $\mathrm{d}\sigma\big(\mathcal{X}_A(t)\big)/\,\mathrm{dt} \,=\, 0$, respectively. An example of the time course of the probability density of an irreversible monomolecular reaction is shown in figure 4.20.

4.3.2.2 Reversible monomolecular reaction

The analysis of the irreversible reaction is readily extended to the reversible case (4.94) that corresponds to a one step birth-and-death process. Again we are dealing with a closed system, the conservation relation $\mathcal{X}_A(t)+\mathcal{X}_B(t) = n_0$ – with n_0 being again the number of molecules of class **A** initially present, $P_n(0) = \delta_{n,n0}$ – holds and the step-up and step-down transition probabilities are given by: $w_n^+ = l\,(n_0 - n)$ and $w_n^- = k\,n$.[46] The master equation is now of the form

[46] Here we note the existence of barriers at $n = 0$ and $n = n_0$, which are characterized by $w_0^- \,=\, 0$, $w_0^+ \,=\, l\,n_0 \,>\, 0$ and $w_{n_0}^+ \,=\, 0$, $w_{n_0}^- \,=\, k\,n_0 \,>\, 0$, respectively. These equations fulfil the conditions for reflecting natural boundaries (section 5.2.1).

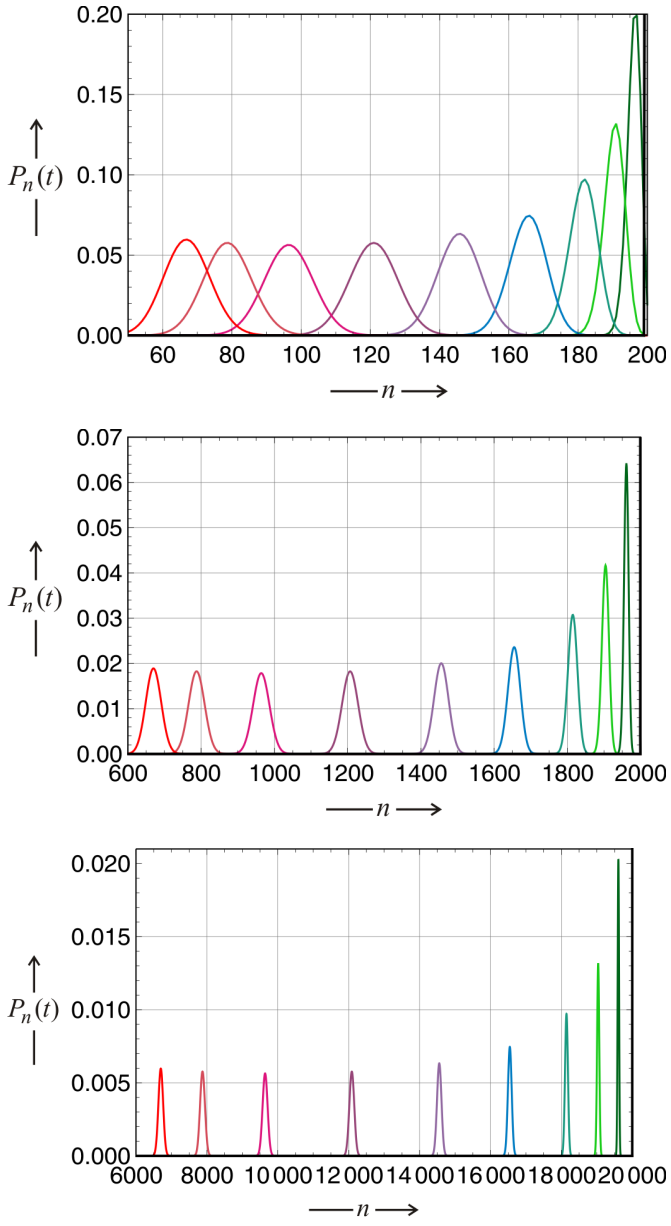


Fig. 4.21 Continued on next page.

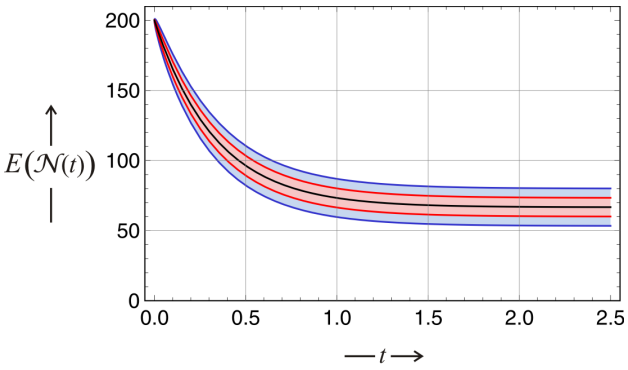


Fig. 4.21 Probability density of a reversible monomolecular reaction The three plots on the previous page show the evolution of the probability density, $P_n(t)$, of the number of molecules of a compound **A** which undergo a reaction **A**⇌**B**. The initially infinitely sharp density $P_n(0) = \delta_{n,n_0}$ becomes broader with time until the variance settles down at the equilibrium value eventually passing a point of maximum variance. On this page we show the expectation value $E\big(\mathcal{X}_A(t)\big)$ and the confidence intervals $E \pm \sigma$ (68,3%,red) and $\pm 2\sigma$ (95,4%,blue) with $\mathrm{var}\big(\mathcal{X}_A(t)\big)$ being the variance. Parameters used: $n_0 = 200$, 2000, and 20 000; $k = 2\,l = 1\,[t^{-1}]$; sampling times: 0 (black), 0.01 (dark green), 0.025 (green), 0.05 (turquoise), 0.1 (blue), 0.175 (blue violet), 0.3 (purple), 0.5 (magenta), 0.8 (deep pink), 2 (red). The initial condition for the time dependence of the expectation value was $n_0 = 200$.

$$\begin{aligned}\frac{\partial P_n(t)}{\partial t} &= l(n_0 - n + 1)P_{n-1}(t) + k(n+1)P_{n+1}(t) - \\ &\quad - \Big(kn + l(n_0 - n)\Big)P_n(t)\ .\end{aligned} \tag{4.101}$$

Making use of the probability generating function $g(s,t)$ we derive the PDE

$$\frac{\partial g(s,t)}{\partial t} = \Big(k + (l-k)s - ks^2\Big)\frac{\partial g(s,t)}{\partial s} + n_0\, l(s-1)\, g(s,t)\ .$$

The solutions of the PDE are simpler when expressed in terms of parameter combinations, $\kappa = \tau_R^{-1} = k + l$ and the equilibrium constant $K = k/l$:

$$\begin{aligned}g(s,t) &= \Big(1 + (s-1)\,e^{-\kappa t} - \frac{s}{K}\Big)^{n_0} = \\ &= \left(\frac{K\,(1 - e^{-\kappa t}) + s\,(Ke^{-\kappa t} + 1)}{1 + K}\right)^{n_0} = \\ &= \sum_{n=0}^{n_0}\left(\binom{n_0}{n}\,\big(Ke^{-\kappa t} + 1\big)^n\,\Big(K(1 - e^{-\kappa t})\Big)\right)^{n_0 - n}\frac{s^n}{(1+K)^{n_0}}\ .\end{aligned}$$

The probability density for the reversible reaction is then obtained as

$$P_n(t) \;=\; \binom{n_0}{n} \frac{1}{(1+K)^{n_0}} \left(Ke^{-\kappa t}+1\right)^n \left(K(1-e^{-\kappa t})\right)^{n_0-n} . \qquad (4.102)$$

Taking the limit to infinite time yields the equilibrium density

$$\lim_{t\to\infty} P_n(t) \;=\; \overline{P}_n \;=\; \binom{n_0}{n} \frac{K^{n_0-n}}{(1+K)^{n_0}} \;=\; \binom{n_0}{n} \frac{1}{(k+l)^{n_0}}\, k^{n_0-n}\, l^n \;, \quad (4.78')$$

which is, of course, identical with the expression (4.78') derived earlier.

Expectation value and variance of the numbers of molecules are readily computed and introducing the function $\omega(t) = K\exp(-\kappa t)+1$ yields:

$$\begin{aligned} E\big(\mathcal{X}_A(t)\big) \;&=\; \frac{n_0}{1+K}\,\omega(t) \;, \\ \operatorname{var}\big(\mathcal{X}_A(t)\big) \;&=\; \frac{n_0\,\omega(t)}{1+K}\left(1-\frac{\omega(t)}{1+K}\right) \;, \end{aligned} \qquad (4.103)$$

and the stationary values are

$$\begin{aligned} \lim_{t\to\infty} E\big(\mathcal{X}_A(t)\big) \;&=\; n_0\,\frac{l}{k+l} \;, \\ \lim_{t\to\infty} \operatorname{var}\big(\mathcal{X}_A(t)\big) \;&=\; n_0\,\frac{k\,l}{(k+l)^2} \;, \\ \lim_{t\to\infty} \sigma\big(\mathcal{X}_A(t)\big) \;&=\; \sqrt{n_0}\,\frac{\sqrt{k\,l}}{k+l} \;. \end{aligned} \qquad (4.104)$$

This result shows that the $\sqrt{N}$-law is fulfilled up to a factor that is independent of N: $E/\sigma = \sqrt{n_0}\,l/\sqrt{k\,l}$. We remark that the deterministic solution

$$a(t) \;=\; \frac{a_0}{1+K}\left(K\,e^{-\kappa t}+1\right)$$

coincides exactly with the expectation value

Starting from a sharp distribution, $P_n(0) = \delta_{n,n0}$, the variance increases, may or may not pass through a maximum and eventually reaches the equilibrium value, $\bar{\sigma^2} = kl\,n_0/(k+l)^2$. The time of maximal fluctuations is easily calculated from the condition $\mathrm{d}\sigma^2/\,\mathrm{dt} = 0$ and one obtains

$$t_{\max\,\mathrm{var}} \;=\; \frac{1}{k+l}\,\ln\left(\frac{2\,k}{k-l}\right) \;. \qquad (4.105)$$

Depending on the sign of $(k-l)$ the approach towards equilibrium passes a maximum value or not. The maximum is readily detected from the height of the mode of $P_n(t)$ as seen in figure 4.21 where a case with $k > h$ is presented.

In order to illustrate fluctuations and their sizes under equilibrium conditions the Austrian physicist Paul Ehrenfest designed a game called Ehrenfest's urn model [104], which was indeed played in order to verify the $\sqrt{N}$-law. Balls, $2N$ in total, are numbered consecutively, $1, 2, \ldots, 2N$, and distributed arbitrarily over two containers, say **A** and **B**. A lottery machine draws lots, which carry the numbers of the balls. When the number of a ball is drawn, the ball is put from one container into the other. This setup is already sufficient for a simulation of the equilibrium condition. The more balls are in a container, the more likely it is that the number of one of its balls is drawn and a transfer occurs into the other container. Just as it occurs with chemical reactions we have self-controlling fluctuations: Whenever a fluctuations becomes large it creates a force for compensation which is proportional to the size of the fluctuation.

4.3.3 Bimolecular chemical reactions

Three classes of bimolecular reactions corresponding to the elementary steps (4.1j), (4.1f), and (4.1g):

$$2\,\mathbf{A} \underset{l}{\overset{k}{\rightleftharpoons}} \mathbf{C}\,, \tag{4.106a}$$

$$\mathbf{A} + \mathbf{B} \underset{l}{\overset{k}{\rightleftharpoons}} \mathbf{C} \quad \text{and} \tag{4.106b}$$

$$\mathbf{A} + \mathbf{X} \underset{l}{\overset{k}{\rightleftharpoons}} 2\,\mathbf{X}\,. \tag{4.106c}$$

are accessible in the irreversible limit, $l \to 0$, to full stochastic analysis [20, 83, 84, 225, 273, 318]. Bimolecularity gives rise to nonlinearities in the kinetic differential equations and in the master equations and complicates substantially the analysis. Bimolecular reactions from the first two classes do not show substantial differences in the qualitative behavior compared to the corresponding monomolecular case $\mathbf{A} \to \mathbf{B}$. The exact coincidence of the expectation value and the deterministic solution, however, in non longer valid. For the solution of the master equations we present here the direct PDE approach as before and compare it to another technique based on Laplace transforms. Autocatalysis in the form of reaction (4.106c) gives rise to intrinsic rate enhancement (section 4.1.1) and different behavior of fluctuations but still reaction dynamics remains simple in the sense of monotonous approach towards unique stationary states. Autocatalytic processes of higher molecularity like, for example, the termolecular step in the Brusselator reaction,

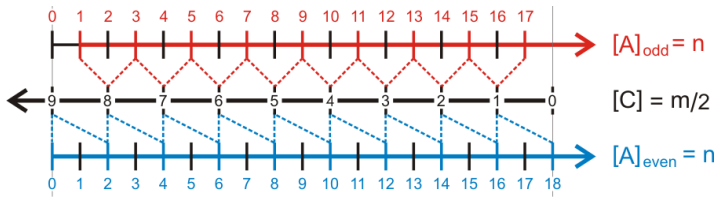


Fig. 4.22 Scaling of particle numbers n and $m/2$ in dimerization reactions. Two molecules **A** react to yield on molecule **C** and accordingly not all molecules **A** can react in case a_0 is odd (red). The other irreducible equivalence class with an even number of **A** molecules (blue) is disjunct. In other words if $\mathcal{X}_{\mathsf{A}}(t_0)$ is odd or even it will remain so for all times t.

$\mathbf{A} + 2\mathbf{X} \rightarrow 3\mathbf{X}$ (4.1m), may give rise to multiple steady states, oscillations of concentrations, and deterministic chaos (section 5.1).

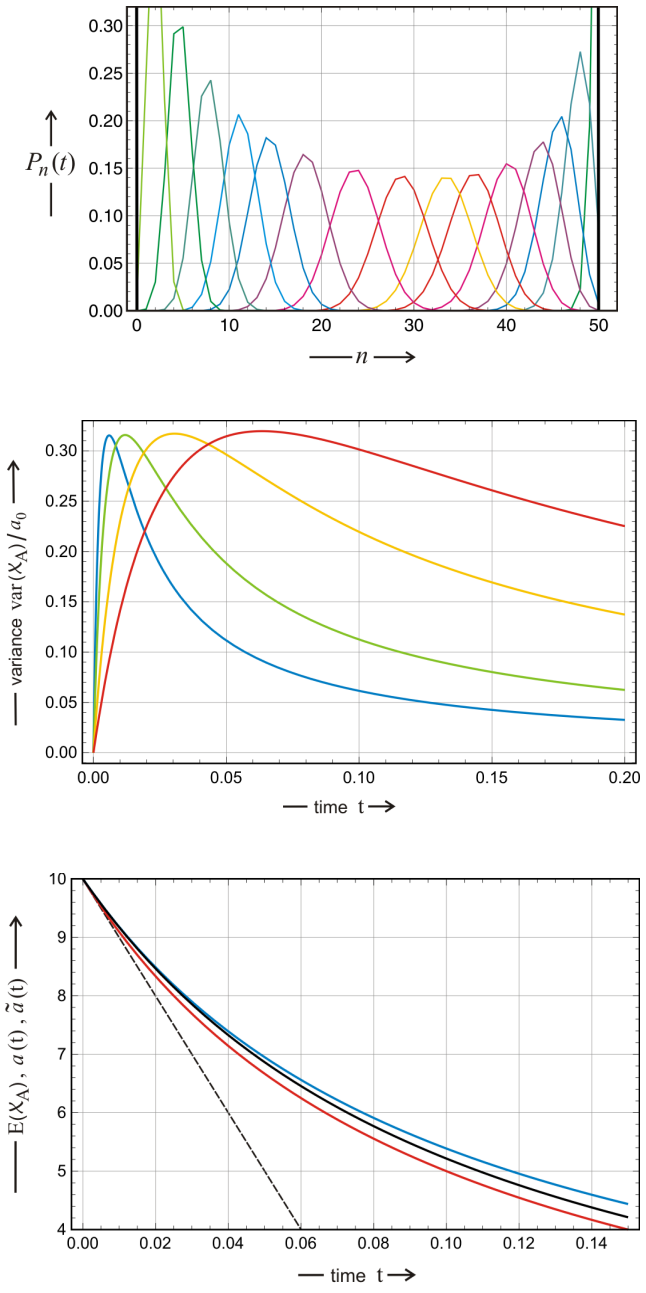


Fig. 4.23 Continued on next page.

Fig. 4.23 Irreversible dimerization reaction $2\mathsf{A} \to \mathsf{C}$. The plot at the top shows the probability distribution $P_n(t) = P\big(\mathcal{X}_\mathsf{C}(t) = n\big)$ describing the number of molecules of species **C** as a function of time and calculated by equation (4.120). The number of molecules **C** is given by the distribution $P_m(t) = P\big(\mathcal{X}_\mathsf{C}(t) = m\big)$. The initial conditions are chosen to be $\mathcal{X}_\mathsf{A}(t) = \delta(n, a_0)$, and $\mathcal{X}_\mathsf{C}(t) = \delta(m, 0)$ and hence we have $n + 2m = a_0$. With increasing time the peak of the distribution moves from right to left. The state $n = 0$ is an absorbing state and hence the long time limit of the system is: $\lim_{t\to\infty} \mathcal{X}_\mathsf{A}(t) = \delta(n, 0)$ $\lim_{t\to\infty} \mathcal{X}_\mathsf{C}(t) = \delta(m, a_0/2)$. Parameters used: $a_0 = 100$ and $k = 0.02\,[t^{-1} \cdot M^{-1}]$; sampling times (upper part): $t = 0$ (black), 0.01 (green), 0.1 (turquoise), 0.2 (blue), 0.3 (violet), 0.5 (magenta), 0.75 (red), 1.0 (yellow), 1.5 (red), 2.25 (magenta), 3.5 (violet), 5.0 (blue), 7.0 (cyan), 11.0 (turquoise), 20.0 (green), 50.0 (chartreuse), and ∞ (black). The plot in the middle shows the variance $\mathrm{var}\big(\mathcal{X}_\mathsf{A}(t)\big)$ as a function of time for different initial conditions: $k = 1$; a_0=10 (blue), 20 (green), 50 (yellow), and 100 (red). The plot at the bottom compares stochastic and deterministic solutions: $\mathrm{E}\big(\mathcal{X}_\mathsf{A}(t)\big)$ (black), $a(t)$ (red) and $\widetilde{a}(t)$ (blue); $k = 1$, $a_0 = 10$. The broken black line shows the asymptotic tangent to all three curves at $t = 0$.

4.3.3.1 Direct solution of PDE

Dimerization reaction: $2\,\mathsf{A} \to \mathsf{C}$. Like the irreversible monomolecular reaction $\mathsf{A} \to \mathsf{B}$ the dimerization reaction (4.106a) is a pure death process and when modeled by means of a master equation [318] we have to take into account that two molecules **A** vanish at a time to form one **C** molecule, and for $\mathcal{X}_\mathsf{A}(t) = n(t)$ an individual jump involves always $\Delta n = 2$. The stoichiometry, which is different from the monomolecular reaction, creates two irreducible equivalence classes (figure 4.15) comprised of odd or even numbers of **A** molecules, respectively (figure 4.22). In other words when the initial number of **A** molecules is odd, $a_0 = 2\mu + 1$ with $\mu \in \mathbb{N}$, $\mathcal{X}_\mathsf{A}$ will always be odd and the last **A** molecule will be unable to react, and for an initially even number of **A** molecules, $a_0 = 2\mu$, $\mathcal{X}_\mathsf{A}$ will always be even and $\mathcal{X}_\mathsf{A} = 0$ is allowed:

$$\mathcal{X}_\mathsf{A}(0) \;=\; a_0 \quad\Longrightarrow\quad \lim_{t\to\infty} \mathcal{X}_\mathsf{A}(t) \;=\; \begin{cases} 1 & \text{if } a_0 = 2\mu + 1 \\ 0 & \text{if } a_0 = 2\mu \end{cases} \qquad \mu \in \mathbb{N}_{\geq 0}\ .$$

The master equation is of the form,

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; \frac{1}{2}\,k\,(n+2)(n+1)\,P_{n+2}(t) \;-\; \frac{1}{2}\,k\,n(n-1)\,P_n(t)\ , \qquad (4.106\text{a'})$$

with $P_n(t) = P\big(\mathcal{X}_\mathsf{A}(t) = n\big)$ with $n \in \mathbb{N}_{\leq 0}$ and $P_n(0) = \delta_{n,n_0}$ with $a_0 = n_0$ and $P\big(\mathcal{X}_\mathsf{C}(0) = c\big) = \delta_{c,0}$ as initial conditions. The master equation gives rise to the following PDE for the probability generating function [318]:

$$\frac{\partial g(s,t)}{\partial t} \;=\; \frac{k}{2}\,(1 - s^2)\,\frac{\partial^2 g(s,t)}{\partial s^2}\ . \qquad (4.107)$$

The analysis of this PDE is more involved than it might appear at a first glance.

For the initial condition $P_n(0) = \delta_{n,n_0}$ and proper boundary conditions exact solutions of equation (4.107) in terms of auxiliary functions are available:

$$g(s,t) \;=\; \sum_{j=0}^{\infty} A_j \, C_j^{-\frac{1}{2}}(s) \, T_j(t) \;, \tag{4.108}$$

wherein the parameters and functions are defined by

$$\begin{aligned} A_j \;&=\; \frac{2j-2}{2^j} \cdot \frac{\Gamma(n_0+1)\,\Gamma\big((n_0-j+1)/2\big)}{\Gamma(n_0-j+1)\,\Gamma\big((n_0+j+1)/2\big)} \;, \\ C_j^{-\frac{1}{2}}(s) \;&:\; (1-s^2)\,\frac{\mathrm{d}^2 C_j^{-\frac{1}{2}}(s)}{\mathrm{ds}^2} \;+\; j(j-1)\,C_j^{-\frac{1}{2}}(s) \;=\; 0 \;, \\ T_j(t) \;&=\; \exp\Big(-\frac{1}{2}\,k\,j(j-1)\,t\Big) \;, \end{aligned}$$

The functions $C_j^{-\frac{1}{2}}(s)$ are ultraspherical or Gegenbauer polynomials named after the German mathematician Leopold Gegenbauer [1, ch.22, pp.773-802]. They are solution of the differential equation shown above and belong to the family of hypergeometric functions. It is straightforward to write down expressions for the expectation values and the variance of the stochastic variable $\mathcal{X}_A(t)$ – μ stands for an integer running index, $\mu \in \mathbb{N}$:

$$\begin{aligned} \mathrm{E}\big(\mathcal{X}_A(t)\big) \;&=\; \sum_{j=2\mu=2}^{2\lfloor \frac{n_0}{2} \rfloor} A_j \, T_j(t) \quad \text{and} \\ \mathrm{var}\big(\mathcal{X}_A(t)\big) \;&=\; \sum_{j=2\mu=2}^{2\lfloor \frac{n_0}{2} \rfloor} \Big(\frac{1}{2}(j^2-j+2)\,A_j\,T_j(t) \;-\; A_j^2\,T_j^2(t)\Big) \;. \end{aligned} \tag{4.109}$$

In order to obtain concrete results these expressions can be readily evaluated numerically. An example of the time course of the probability density function is shown in figure 4.23 (For a comparison of the relative widths of the densities of all three bimolecular reactions see figure 4.28).

There is one interesting detail in the deterministic version of the dimerization reaction. With $[\mathbf{A}]{=}a(t)$ and $a(0) = a_0$, it is conventionally modeled by the differential equation (4.110a) for which an exact analytical solution is readily derived. Although particle numbers in chemical reactions are commonly so large that a^2 is indistinguishable from $a\,(a{-}1)$ for practical purposes it is worth considering the corrected kinetic equation (4.110b) for which we have also an exact solution:

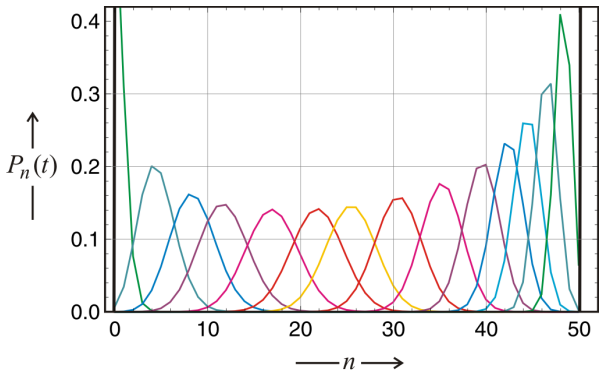


Fig. 4.24 Irreversible association reaction $\mathbf{A} + \mathbf{B} \to \mathbf{C}$. The plot shows the probability distribution $P_n(t) = P\big(\mathcal{X}_\mathsf{C}(t) = n\big)$ describing the number of molecules of species **C** as a function of time and calculated by equation (4.119). The initial conditions are chosen to be $\mathcal{X}_\mathsf{A}(t) = \delta(a, a_0)$, $\mathcal{X}_\mathsf{B}(t) = \delta(b, b_0)$, and $\mathcal{X}_\mathsf{C}(t) = \delta(c, 0)$. With increasing time the peak of the distribution moves from left to right. The state $n = \min(a_0, b_0)$ is an absorbing state and hence the long time limit of the system is: $\lim_{t\to\infty} \mathcal{X}_\mathsf{C}(t) = \delta\big(n, \min(a_0, b_0)\big)$. Parameters used: $a_0 = 50$, $b_0 = 51$, $k = 0.02\,[t^{-1} \cdot M^{-1}]$; sampling times (upper part): $t = 0$ (black), 0.01 (green), 0.1 (turquoise), 0.2 (blue), 0.3 (violet), 0.5 (magenta), 0.75 (red), 1.0 (yellow), 1.5 (red), 2.25 (magenta), 3.5 (violet), 5.0 (blue), 7.0 (cyan), 11.0 (turquoise), 20.0 (green), and ∞ (black).

$$-\frac{\mathrm{d}a}{\mathrm{d}t} = \frac{1}{2}\,k\,a^2 \quad \Longrightarrow \quad a(t) = \frac{a_0}{1 + a_0\,k\,t/2} \quad \text{and} \tag{4.110a}$$

$$-\frac{\mathrm{d}\widetilde{a}}{\mathrm{d}t} = \frac{1}{2}\,k\,\widetilde{a}(\widetilde{a} - 1) \quad \Longrightarrow \quad \widetilde{a}(t) = \frac{\widetilde{a}_0}{\widetilde{a}_0 + (1 - \widetilde{a}_0)e^{-kt/2}}\,. \tag{4.110b}$$

At not too long times the expectation value of the stochastic solution lies always between the two solution curves (4.110a) and (4.110b). In the limit $t \to 0$ all three curves converge to the asymptotic limit for small times: $\widehat{a}(t) = a_0(1 - a_0\,k\,t/2)$. At short times the expectation value $\mathrm{E}\big(\mathcal{X}_\mathsf{A}(t)\big)$ and the corrected deterministic curve $\widetilde{a}(t)$ come closer together whereas the expectation value comes closer to the conventional deterministic curve $a(t)$. At large times the conventional deterministic curve and the stochastic curve may even cross.[47]

Association reaction: $\mathbf{A} + \mathbf{B} \to \mathbf{C}$. In the second example, the association reaction (4.106b), we are dealing with three dependent stochastic variables $\mathcal{X}_\mathsf{A}(t)$, $\mathcal{X}_\mathsf{B}(t)$, and $\mathcal{X}_\mathsf{C}(t)$. Following Donald McQuarrie and coworkers [318] we define the probability $P_n(t) = P\big(\mathcal{X}_\mathsf{A}(t) = n\big)$ and apply the standard

[47] This result is an artifact of the unability of the continuous deterministic function to distinguish between odd and even numbers of molecules.

initial condition $P_n(0) = \delta_{n,n_0}$ with $n_0 = a_0$, $P(\mathcal{X}_{\mathsf{B}}(0) = b) = \delta_{b,b_0}$, and $P(\mathcal{X}_{\mathsf{C}(0)} = c) = \delta_{c,0}$. From the laws of stoichiometry and mass conservation we have $\mathcal{X}_{\mathsf{B}}(t) = b_0 - n_0 + \mathcal{X}_{\mathsf{A}}(t)$ and $\mathcal{X}_{\mathsf{C}}(t) = n_0 - \mathcal{X}_{\mathsf{A}}(t)$. For simplicity we denote $b_0 - a_0 = \vartheta_0$. Then the master equation for the chemical reaction is of the form

$$\frac{\partial P_n(t)}{\partial t} \;=\; k\,(n+1)\,(\vartheta_0 + n + 1)\,P_{n+1}(t) \;-\; k\,n\,(\vartheta_0 + n)\,P_n(t)\;. \quad (4.106\text{b'})$$

We remark that the step-down transition probabilities are no longer linear in n. The corresponding PDE for the generating function is readily calculated

$$\frac{\partial g(s,t)}{\partial t} \;=\; k\,(\vartheta_0+1)(1-s)\frac{\partial g(s,t)}{\partial s} \;+ k\,s(1-s)\,\frac{\partial^2 g(s,t)}{\partial s^2}\;. \qquad (4.111)$$

The derivation of solutions for this PDE is quite demanding, but as in case of the dimerization reaction it can be achieved by separation of variables:

$$g(s,t) \;=\; \sum_{m=0}^{\infty} A_m\,Z_m(s)\,T_m(t)\;. \qquad (4.112)$$

We list here only the coefficients and functions of the solution:

$$\begin{aligned} A_m \;&=\; (-1)^m\,\frac{(2m+\vartheta_0)\Gamma(m+\vartheta_0)\Gamma(n_0+1)\Gamma(n_0+\vartheta_0+1)}{\Gamma(m+1)\Gamma(\vartheta_0+1)\Gamma(n_0-m+1)\Gamma(n_0+\vartheta_0+m+1)}\;,\\ Z_m(s) \;&=\; J_m(\vartheta_0,\vartheta_0+1,s)\;,\quad \text{and}\\ T_m(t) \;&=\; \exp\bigl(-m(m+\vartheta_0)\,k\,t\bigr)\;. \end{aligned}$$

Herein, Γ represents the conventional *gamma function* with the definition $\Gamma(x+1) = x\,\Gamma(x)$, and $J(p,q,s)$ are the Jacobi polynomials named after the German mathematician Carl Jacobi [1, ch.22, pp.773-802], which are solutions of the differential equation

$$s(1-s)\frac{\mathrm{d}^2 J_n(p,q,s)}{\mathrm{ds}^2} \;+\; \Bigl(q-(p+1)s\Bigr)\frac{\mathrm{d}J_n(p,q,s)}{\mathrm{ds}} \;+\; n(n+p)\,J_n(p,q,s) \;=\; 0\;.$$

These polynomials fulfil the following conditions:

$$\begin{aligned} &\frac{\mathrm{d}J_n(p,q,s)}{\mathrm{ds}} \;=\; -\,\frac{n(n+p)}{s}\,J_{n-1}(p+2,q+1,s) \quad \text{and}\\ &\int_0^1 s^{q-1}(1-s)^{p-q}J_n(p,q,s)\,J_\ell(p,q,s)\,ds \;=\; \frac{n!\,\bigl(\Gamma(q)\bigr)^2\Gamma(n+p-q+1)}{(2n+p\,)\Gamma(n+p)\Gamma(n+q)}\,\delta_{\ell,n}\;. \end{aligned}$$

We differentiate twice at the value $s = 1$ of the dummy variable and find:

$$\left(\frac{\partial g(s,t)}{\partial s}\right)_{s=1} = \sum_{m=1}^{n_0} \frac{(2m+\vartheta_0)\Gamma(n_0+1)\Gamma(n_0+\vartheta_0+1)}{\Gamma(n_0-m+1)\Gamma(n_0+\vartheta_0+m+1)}\, T_m(t)\,, \tag{4.113}$$

$$\left(\frac{\partial^2 g(s,t)}{\partial s^2}\right)_{s=1} =$$

$$= \sum_{m=2}^{n_0} \frac{(m-1)(m+\vartheta_0+1)(2m+\vartheta_0)\Gamma(n_0+1)\Gamma(n_0+\vartheta_0+1)}{\Gamma(n_0-m+1)\Gamma(n_0-\vartheta_0+m+1)}\, T_m(t) \tag{4.114}$$

from which we obtain expectation value and variance

$$E(\mathcal{X}_A(t)) = \left(\frac{\partial g(s,t)}{\partial s}\right)_{s=1} \quad \text{and}$$

$$\mathrm{var}(\mathcal{X}_A(t)) = \left(\frac{\partial^2 g(s,t)}{\partial s^2}\right)_{s=1} + \left(\frac{\partial g(s,t)}{\partial s}\right)_{s=1} - \left(\left(\frac{\partial g(s,t)}{\partial s}\right)_{s=1}\right)^2 . \tag{2.25'}$$

As we see in the current example and we shall see again in the next section, bimolecularity complicates the solution of the chemical master equations substantially and makes solutions quite sophisticated. We dispense here from the detailed expressions but provide the results for the special case of vast excess of one reaction partner, $|\vartheta_0| \gg n_0 > 1$, which is known as *pseudo first order condition* or *concentration buffering*. Then, the sums can be approximated well be the first terms and we find with $k' = \vartheta_0 k$:

$$\left(\frac{\partial g(s,t)}{\partial s}\right)_{s=1} \approx n_0 \frac{\vartheta_0+2}{n_0+\vartheta_0+1}\, e^{-(\vartheta_0+1)kt} \approx n_0\, e^{-k't} \quad \text{and}$$

$$\left(\frac{\partial^2 g(s,t)}{\partial s^2}\right)_{s=1} \approx n_0\,(n_0-1)\, e^{-2\,k't}\,,$$

and we obtain finally,

$$\begin{aligned} E\big(\mathcal{X}_A(t)\big) &= n_0\, e^{-k't} \quad \text{and} \\ \mathrm{var}\big(\mathcal{X}_A(t)\big) &= n_0\, e^{-k't}\left(1 - e^{-k't}\right)\,, \end{aligned} \tag{4.115}$$

which is formally the same result as obtained for the irreversible first order reaction with $k \Rightarrow k' = \vartheta_0 k$.

4.3.3.2 Laplace transform of master equations

Because of its more general applicability we consider here also the solution of chemical master equations by means of Laplace transform [20, 225, 273], which is similar to the approach used in section 3.2.4 for analyzing random walks. The probability density of a master equation, $P_m(t)$, is Laplace transformed

$$V_m(s) = \int_0^\infty \exp(-s\,t)\, P_m(t)\,\mathrm{dt}\,,$$

and the time derivative is calculated through integration by parts:

$$\int_0^\infty \frac{\mathrm{d}P_m(t)}{\mathrm{dt}}\, e^{-s\,t}\,\mathrm{dt} \;=\; P_m(t)\, e^{-s\,t}\,\Big|_0^\infty - \int_0^\infty P_m(t)\, e^{-s\,t}(-s)\,\mathrm{dt} \;=$$

$$= \; s\int_0^\infty P_m(t)\, e^{-s\,t}\,\mathrm{dt} - P_m(0) \;=\; s\,V_m(s) - P_m(0)\;.$$

Thereby we obtain an algebraic equation for the Laplace transform $V_m(s)$, which can be solved by a standard technique and then the probability density in obtained through backtransformation by inverse Laplace transform.

The inverse Laplace transformation by Mellin's fomula is defined by[48]

$$f(t) \;=\; \mathcal{L}^{-1}\Big(F(s)\Big) \;\doteq\; \frac{1}{2\pi\dot{\imath}} \lim_{\vartheta\to\infty} \int_{\gamma-\dot{\imath}\vartheta}^{\gamma+\dot{\imath}\vartheta} e^{s\,t} F(s)\,\mathrm{d}s\;. \tag{4.116}$$

where γ is a real number chosen such that the contour path of integration is the region of convergence of $F(s)$. If the integral is extended over the positive real axis, all poles of the function have to lie left to the imaginary axis. For illustration we repeat first the transformation properties of the simple exponential function $f(t) = \exp(-a\,t)$:

$$\mathcal{L}\big(e^{-a\,t}\big) \;=\; \frac{1}{s+a} \;\;\Longrightarrow\;\; \mathcal{L}^{-1}\Big(\frac{1}{s+a}\Big) \;=\; e^{-a\,t} \tag{4.117}$$

Here, as required the pole is indeed situated left to the imaginary axis.

Irreversible association reaction: $\mathbf{A} + \mathbf{B} \to \mathbf{C}$. We write down the master equation in a slightly different form and introduce $\mathcal{X}_\mathsf{C}(t)$ as the stochastic variable counting the number of molecules **C** in the system: $P_m(t) = P\big(\mathcal{X}_\mathsf{C}(t) = m\big)$. With the initial condition $P_m(0) = \delta_{m,0}$ and the upper limit of m, $\lim_{t\to\infty} P_m(t) = \gamma$ with $\gamma = \min\{a_0, b_0\}$ where a_0 and b_0 are the sharply defined numbers of **A** and **B** molecules initially present, $P\big(\mathcal{X}_\mathsf{A}(0) = a\big) = \delta_{a,a_0}$, $P\big(\mathcal{X}_\mathsf{B}(0) = b\big) = \delta_{b,b_0}$, we have

$$\sum_{m=0}^{\gamma} P_m(t) = 1\,,\;\; m \in \mathbb{Z}, \quad \text{and thus} \quad P_m(t) = 0 \;\forall\; m \notin [0,\gamma]$$

and the master equation now takes on the form

[48] Mellin's formula is named after the Finnish mathematician Hjalmar Mellin and represents one of several possible definitions. Another frequently used expression is Post's inversion formula of the Polish-American mathematician Emil Leon Post, which develops the inversion into a derivative of infinite order.

$$\begin{aligned}\frac{\mathrm{d}P_m(t)}{\mathrm{dt}} &= k\,\big(a_0-(m-1)\big)\big(b_0-(m-1)\big)P_{m-1}(t) - \\ &\quad - k\,(a_0-m)(b_0-m)P_m(t)\,,\ 0\le m\le\gamma\ . \qquad (4.106\text{b''})\end{aligned}$$

Laplace transform and insertion of $P_m(0)=\delta(m,0)$ leads to

$$\begin{aligned}-1+s\,V_0(s) &= -k\,a_0\,b_0\,V_0(s)\ ,\\ s\,V_m(s) &= k\,\big(a_0-(m-1)\big)\big(b_0-(m-1)\big)\,V_{m-1}(s) - \\ &\quad - k\,(a_0-m)(b_0-m)\,V_m(s)\,,\ 1\le m\le\gamma\ ,\ \text{and}\\ s\,V_\gamma(s) &= k\,\big(a_0-(\gamma-1)\big)\big(b_0-(\gamma-1)\big)\,V_{\gamma-1}(s)\ .\end{aligned}$$

The solutions in Laplace space expressed by the functions $V_m(s)$ are calculated by successive iteration, $V_0(s)\to V_1(s)\to\cdots\to V_\gamma(s)$, that yields

$$V_m(s) = \binom{a_0}{m}\binom{b_0}{m}(m!)^2k^n\prod_{j=0}^{n}\frac{1}{s+k(a_0-j)(b_0-j)}\,,\ 0\le m\le\gamma\ , \qquad (4.118\text{a})$$

where the product in the denominator is resolved into partial fractions

$$\begin{aligned}\prod_{j=0}^{n}\Big(s+k(a_0-j)(b_0-j)\Big)^{-1} &= \sum_{j=0}^{n}\frac{A_j}{s+k(a_0-j)(b_0-j)}\ \text{ with}\\ A_j &= (-1)^{m+j}k^{-m}\,\frac{(a_0+b_0-2j)\,(a_0+b_0-m-j-1)!}{j!(m-j)!(a_0+b_0-j)!}\ ,\end{aligned} \qquad (4.118\text{b})$$

which are suitable for inverse Laplace transform.

Exchanging integration and summation for the inverse transformation and performing the transform yields the final result

$$\begin{aligned}P_m(t) &= (-1)^m\binom{a_0}{m}\binom{b_0}{m}\sum_{j=0}^{m}(-1)^j\left(1+\frac{m-j}{a_0+b_0-m-j}\right)\times\\ &\quad\times\binom{m}{j}\binom{a_0+b_0-j}{m}^{-1}e^{-k(a_0-j)(b_0-j)t}\ .\end{aligned} \qquad (4.119)$$

An illustrative example is shown in figure 4.24. The difference between the two irreversible reactions, monomolecular conversion and bimolecular association (figure 4.20), is indeed not spectacular.

Dimerization reaction: $2\mathbf{A} \rightarrow \mathbf{C}$. The master equation of the dimerization reaction has been solved by means of a Laplace transform [225] in full analogy to the procedure described in the previous paragraph deling with the association reaction. For this goal we rewrite down the master equation in a slightly different form that takes into account that $\Delta\mathcal{X}_{\mathsf{A}} = 2$:

$$\frac{\partial P_{2w}(t)}{\partial t} \;=\; -\frac{1}{2}\,k\,(2w)(2w-1)P_{2w}(t) \;+\; \frac{1}{2}\,k\,(2w+2)(2w+1)P_{2w+2}(t) \quad \text{(4.106a")}$$

with $w \in \mathbb{Z}$ and $P_w(t) = 0 \,\forall\, \{w < 0 \vee w > \gamma \vee 2w = 2\mu + 1 \text{ with } \mu \in \mathbb{N}_{\geq 0}\}$,
being the condition that all probabilities outside the interval $[0, 2\gamma]$ as well as the odd probabilities for odd values of w, $P_{2\mu+1}$ vanish (figure 4.22). Herein $\gamma = \lfloor \frac{a_0}{2} \rfloor$ denotes the maximal number of $\mathbf{C}$ molecules that can be formed. The probability distribution $P_{2w}(t)$ is derived by means of the Laplace transform

$$V_w(s) \;=\; \int_0^\infty \exp\left(-\,s\cdot t\right) P_w(t)\;\mathrm{dt}$$

yielding the set of difference equations

$$\begin{aligned} -1 \;+\; s\,q_{2a_0}(s) \;&=\; -\frac{1}{2}\,k\,(2a_0)(2a_0-1)\,q_{2a_0}(s)\,, \\ s\,q_{2n}(s) \;&=\; -\frac{1}{2}\,k\,(2n)(2n-1)\,q_{2n}(s) + \\ &\quad + \frac{1}{2}\,k\,(2n+2)(2n+1)\,q_{2n+2}(s)\,,\; 0 \leq y \leq a_0 - 1\;, \end{aligned}$$

which again can be solved by successive iteration. It is straightforward to calculate the Laplace transform for 2μ, the number of molecules of species $\mathbf{A}$ that have reacted to yield $\mathbf{C}$: $2\mu = 2(a_0 - m)$ with $m = [\mathbf{C}]$ and $0 \leq m \leq a_0$:

$$q_{2(a_0-m)}(s) \;=\; \left(\frac{k}{2}\right)^m \binom{2a_0}{2m}(2m)! \prod_{j=1}^{m} \left(s + \frac{k}{2}\big(2(a_0-j)\big)\cdot\big(2(a_0-j)-1\big)\right)^{-1}\;,$$

and a somewhat tedious but straightforward exercise in algebra yields the inverse Laplace transform:

$$\begin{aligned} P_{2(a_0-m)}(t) \;&=\; (-1)^m\,\frac{a_0!\,(2a_0-1)!!}{(a-m)!\,(2a_0-2m-1)} \times \\ &\quad \times \sum_{j=0}^{m} (-1)^j \frac{(4a_0-4j-1)(4a_0-2m-2j-3)!!}{j!(m-j)!(4a_0-2j-1)!!} \times \\ &\quad \times\; e^{-k\,(a_0-j)\cdot\left(2(a_0-j)-1\right)t}\;. \end{aligned}$$

The substitution $i = a_0 - j$ leads to

$$
\begin{aligned}
P_{2(a_0-m)}(t) \;&=\; (-1)^m\,\frac{a_0!\,(2a_0-1)!!}{(a-m)!\,(2a_0-2m-1)}\times \\
&\times\;\sum_{i=a_0-m}^{a_0}(-1)^{a_0-i}\frac{(4i-1)(2a_0-2m+2i-3)!!}{(a_0-i)!(a_0-i+m)!(2a_0+2i-1)!!}\times \\
&\times\; e^{-k\,2i\cdot\left(2i-1\right)\,t}\;.
\end{aligned}
$$

Setting now $n = a_0 - m$ in accord with the definition of m we obtain the final result

$$
\begin{aligned}
P_{2n}(t) \;&=\; (-1)^n\,\frac{a_0!(2a_0-1)!!}{n!(2n-1)!!}\times \\
&\times\;\sum_{i=1}^{n}(-1)^i\,\frac{(4i-1)(2n+2i-3)!!}{n!(2n-1)!!}\;\times\; e^{-k\,i(2i-1)t}\;.
\end{aligned}
\tag{4.120}
$$

The results are illustrated be means of a numerical example in figure 4.23.

Reversible association reaction: $\mathbf{A}+\mathbf{B}\rightleftharpoons\mathbf{C}$. The reversible association reaction is described by the chemical master equation

$$
\begin{aligned}
\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;&=\; k\,(n+1)\,(\vartheta_0+n+1)\,P_{n+1}(t)\;-\;k\,n\,(\vartheta_0+n)\,P_n(t)+ \\
&+\;l\,(n_0-n+1)\,P_{n-1}(t)\;-\;l\,(n_0-n)\,P_n(t)
\end{aligned}
\tag{4.121}
$$

and has been solved by application of linear algebra to the probability density vector $\mathbf{P}(t) = \big(P_n(t); n\in\mathbb{N}, n\in[0,a_0]\big)^{\mathrm{t}}$ with $a_0 = n_0$ for $b_0 > a_0$,[49] $P_n(t) = P\big(\mathcal{X}_{\mathsf{A}}(t)=n\big)$, $P_n(0)=\delta_{n,n_0}$, Laplace transform and its inversion [273]. First we write the master equation in terms of step-up and step-down transition probabilities (3.96) whereby for convenience we make a modification of the time axis $k\,t \Rightarrow t$:

$$
w_n^+ \;=\; K\,(a_0-n) \quad\text{and}\quad w_n^- \;=\; n\,(\vartheta_0+n)\;, \tag{4.122}
$$

with $K = l/k$ being the dissociation constant, and obtain the general form

$$
\begin{aligned}
&\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;= \\
&=\begin{cases}
-(w_0^+ + w_0^-)P_0(t) + w_1^- P_1(t)\,, & \text{if } n=0\,,\\
+w_{n-1}^+P_{n-1}(t) - (w_n^+ + w_n^-)P_n(\tau) + w_{n+1}^-P_{n+1}(t)\,, & \text{if } 0<n<n_0\,,\\
+w_{n-2}^+P_{n-2}(t) - (w_{n-1}^+ + w_{n-1}^-)P_n(t)\,, & \text{if } n=n_0\,,
\end{cases}
\end{aligned}
$$

which in vector format reads

[49] If $a_0 > b_0$ we can simply exchange the variables $\mathcal{X}_A(t)$ and $\mathcal{X}_B(t)$ without loosing generality.

$$\frac{\mathrm{d}\mathbf{P}(t)}{\mathrm{dt}} \;=\; \mathrm{W}\cdot\mathbf{P}(t)\;, \tag{4.123}$$

where W is the general tridiagonal transition matrix of a birth-and-death process

$$\mathrm{W} = \begin{pmatrix} -(w_0^+ + w_0^-) & w_1^- & 0 & \dots & 0 & 0 \\ w_0^+ & -(w_1^+ + w_1^-) & w_2^- & \dots & 0 & 0 \\ 0 & w_1^+ & -(w_2^+ + w_2^-) & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -(w_{n_0-2}^+ + w_{n_0-2}^-) & w_{n_0-1}^- \\ 0 & 0 & 0 & \dots & w_{n_0-2}^+ & -(w_{n_0-1}^+ + w_{n_0-1}^-) \end{pmatrix}.$$

Laplace transform of the probability density,

$$V_n(s) \;=\; \int_0^\infty \exp\left(-s\,t\right) P_n(t)\;\mathrm{dt}\;, \tag{4.124}$$

yields an algebraic equation for $\mathbf{V}(s) = \left(V_n(s); n\in\mathbb{N}, n\in[0,a_0]\right)^{\mathrm{t}}$:

$$s\,\mathbf{V}(s) \;=\; \mathrm{W}\cdot\mathbf{V}(s) + \mathbf{P}_0 \;\;\text{and}\;\; \mathbf{V}(s)\left(s\,\mathbb{I} - \mathrm{W}\right) \;=\; \mathbf{P}_0\;, \tag{4.125}$$

where $\mathbf{P}_0 = \left(P_0(0), P_1(0), \dots, P_{a_0}(0)\right)^{\mathrm{t}} = (0,0,\dots,1)^{\mathrm{t}}$. The formal solution of this equation is

$$\mathbf{V}(s) \;=\; \left(s\,\mathbb{I} - \mathrm{W}\right)^{-1}\cdot\mathbf{P}_0\;, \tag{4.126}$$

where matrix inversion is performed in the conventional way

$$\left(s\,\mathbb{I} - \mathrm{W}\right)^{-1} \;=\; \frac{1}{\left|\left(s\,\mathbb{I} - \mathrm{W}\right)\right|}\;\mathrm{adj}\left(s\,\mathbb{I} - \mathrm{W}\right)$$

with 'adj' denoting the *adjugate* or *classical adjoint* matrix.[50] The simple form of $\mathbf{P}_0$ makes it possible to obtain $\mathbf{V}$ using only the elements of the last column of the matrix $\left(s\,\mathbb{I} - \mathrm{W}\right)^{-1}$. After some calculations one obtains the solution in Laplace space [273]:

$$V_n(s) \;=\; \frac{1}{(\vartheta_0+1)_n\,n!}\,\frac{a_0!\,b_0!}{\vartheta_0!}\,\frac{D_n(s)}{D_{a_0-1}(s)}\;, \tag{4.127}$$

where the polynomials $D_n(s)$ can be constructed recursively:

$$D_n(s) \;-\; (s + w_{n-1}^+ + w_{n-1}^-)D_{n-1}(s) + w_{n-2}^+\,w_{n-1}^-\,D_{n-2}(s)\;, \tag{4.128}$$

with $D_0 = 1$ and $D_1 = s + w_0^+$. This recursion, of course, is just an alternative to (4.118) way to calculate the solution in Laplace space.

[50] The adjugate matrix of a square matrix is the transpose of the cofactor matrix (For details see textbooks of linear algebra, e.g., [416, pp. 231-232]).

The final step is the inverse Laplace transformation, which can be done by applying Mellin's formula (4.116) and integration according to the residue theorem [14, p. 444]:

$$P_n(t) \;=\; \sum_{j=0}^{a_0} \lim_{s\to\lambda_j} \Big((s-\lambda_j) \exp{(s\,t)}\, V_n(s) \Big) \;,$$

where the values λ_j are the eigenvalues of the transition matrix W. Combining both results yields the final solution:

$$P_n(t) \;=\; \frac{a_0!\, b_0!}{(\vartheta_0+1)_n\, n!\, (\vartheta_0)!} \sum_{j=0}^{a_0} \frac{D_n(\lambda_j)\, \exp{(\lambda_j t)}}{\Big(\frac{\partial D_{a_0+1}(s)}{\partial s} \Big)\Big|_{s=\lambda_j}} \;, \tag{4.129}$$

In principle,the exact probability density can be calculated from equation (4.129) provided the eigenvalues of matrix W are known. In the general case W is a tridiagonal matrix and the eigenvalues can be obtained only by numerical computation. In some special cases, nevertheless, analytical solutions can be obtained. We mention two examples (i) the irreversible reaction $\mathbf{A}+\mathbf{B} \to \mathbf{C}$ and (ii) the stationary or equilibrium density $\overline{P}_n = \lim_{t\to\infty} P_n(t)$ for the reversible reaction $\mathbf{A} + \mathbf{B} \rightleftharpoons \mathbf{C}$.

For the irreversible reaction the eigenvalues are identical with the diagonal elements of matrix W, which is upper-triangular[51] the eigenvalues coincide with the diagonal elements in this case:

$$w_j^+ \;=\; 0 \;\forall\; j \in \mathbb{N}\,, j \in [0, a_0] \quad \Rightarrow \quad \lambda_j \;=\; w_j^- \;=\; -j\,(\vartheta_0 + j) \;.$$

The expression for the probability density [273] then becomes:

$$\begin{aligned} P_n(t) \;&=\; \frac{a_0!\, b_0!}{(\vartheta_0+n)!} \times \\ &\times \sum_{j=n}^{a_0} (-1)^{j-n} \frac{(\vartheta_0+2j)\,(\vartheta_0+n+j-1)!}{(a_0-j)!\,(j-n)!\,(b_0+j)!}\, e^{-j\,(\vartheta_0+j)kt} \;, \end{aligned} \tag{4.119'}$$

where we have restored the original time axis, $t \Rightarrow k\,t$. Equations (4.119) and (4.119') yield exactly the same density and are mathematically equivalent, although the expressions are different, since m in (4.119) counts the molecules $\mathbf{C}$ whereas n counts molecules $\mathbf{A}$.

The equilibrium probability density may be calculated by making advantage of an interesting relation between a function and its Laplace transform known as Laplace initial and final value theorem:

---

[51] A matrix that has no nonzero entries below the main diagonal is called upper-triangular, and a lower-triangular matrix has only zero elements above the diagonal.

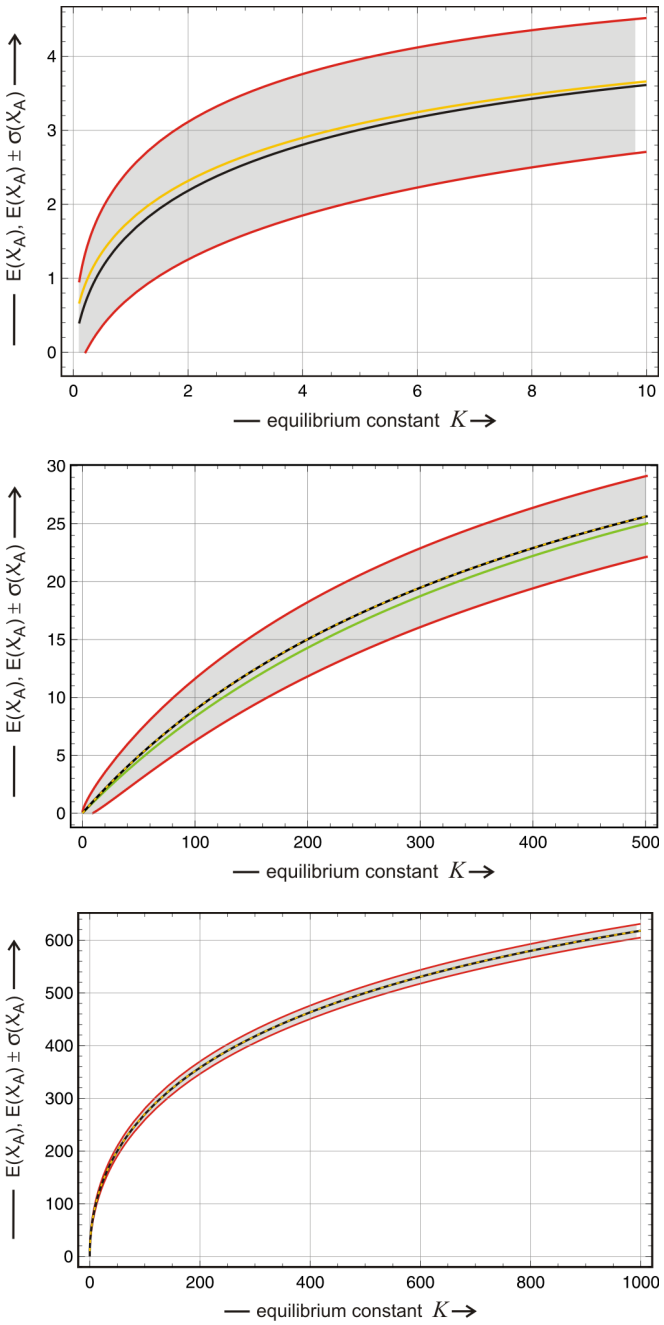


Fig. 4.25 Continued on next page.

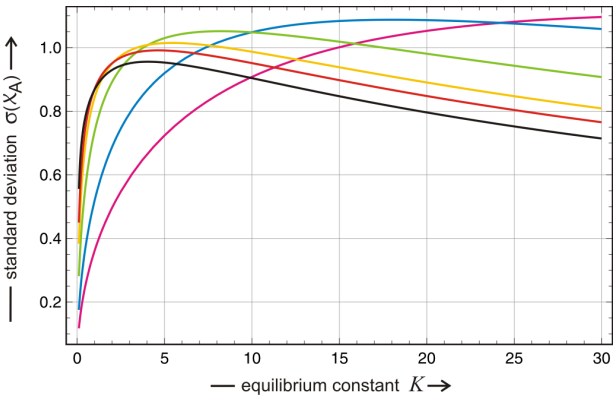


Fig. 4.25 Equilibrium of the association reaction. The first three figures show the equilibrium expectation values $\mathrm{E}\big(\overline{\mathcal{X}}_{\mathsf{A}}(K)\big)$ (black) embedded in the one standard deviation zone $\mathrm{E}(\overline{\mathcal{X}}_{\mathsf{A}}) \pm \sigma(\overline{\mathcal{X}}_{\mathsf{A}})$ (gray with red borders), in addition the deterministic solution $\bar{a}(K)$ (yellow) and the pseudo first order result $\widetilde{a}(K)$ (green) are shown. Choice of parameters: $a_0 = 5, b_0 = 5$ (upper plot), $a_0 = 50, b_0 = 500$ (middle plot), and $a_0 = 1000, b_0 = 1000$ (lower plot). The fourth plot shows the standard deviation as a function of the equilibrium constants, $\sigma\big(\overline{\mathcal{X}}_{\mathsf{A}}(K)\big)$. These curves start from $\sigma = 0$ for $K = 0$, pass a maximum and approach zero in the limit $K \to \infty$. Choice of parameters: $a_0 = 5$, $b_0 = 5$ (black), 6 (red), 7 (yellow), 10 (green), 20 (blue), and 40 (magenta).

$$\lim_{s\to 0} s\,V_n(s) \;=\; P_n(\infty) \;=\; \overline{P}_n \quad \text{and} \quad \lim_{s\to\infty} s\,V_n(s) \;=\; P_n(0)\;. \tag{4.130}$$

In order to calculate the equilibrium density we need to know only the limiting value of the Laplace transformed probability density

$$\lim_{s\to 0} s\,V_n(s) \;=\; \lim_{s\to 0} s\,\frac{a_0!\,b_0!}{(\vartheta_0+n)!\,n!}\,\frac{D_n(s)}{D_{a_0+1}(s)}\;,$$

in particular we require the constant terms of the polynomials $D_n(s)$, which can be obtained from the recursion (4.128):

$$D_n(0) \;=\; \prod_{j=0}^{n-1} w_j^+ \;=\; K^n\,\frac{a_0!}{(a_0-n)!} \;=\; \frac{(a_0-n+1)_n}{K_b^n}\;,$$

where $(\cdot)_n$ is the rising Pochhammer symbol and $K_b = K^{-1} = k/l$ the association or binding constant. From $w_j^+ = (a_0 - j)K$ follows $D_n(0) > 0\,\forall\, n \in [0, a_0]$ and $D_{a_0+1} = 0$ immediately. After some calculation and consideration of the normalization condition $\sum_{n=0}^{a_0} P_j(t) = 1$ we obtain the final result for the stationary distribution

$$\overline{P}_n \;=\; \frac{(a_0-n+1)_n}{K^{-n}\,(\vartheta_0+1)_n\,n!}\cdot\frac{1}{{}_1F_1(-a_0;\vartheta_0+1;-K)}\;, \tag{4.131}$$

where ${}_1F_1(\alpha;\gamma;x)$ is the confluent hypergeometric function:

$${}_1F_1(\alpha;\gamma;x) \;=\; \sum_{j=0}^{\infty}\frac{(\alpha)_j}{(\gamma)_j}\,\frac{x^j}{j!}\;.$$

It is worth recalling that the result for the equilibrium density, $\overline{P}_{\mathbf{n}}$, of the association reaction can be derived much easier form equation (4.77) through elimination of the linear dependencies, $n_{\mathsf{A}}=n$, $n_{\mathsf{B}}=\vartheta_0+n$, and $n_{\mathsf{C}}=a_0-n$:

$$\begin{aligned}\overline{p}_{\mathbf{n}} \;&=\; \frac{\bar{x}_{\mathsf{A}}^{n_{\mathsf{A}}}}{n_{\mathsf{A}}!}\,\frac{\bar{x}_{\mathsf{B}}^{n_{\mathsf{B}}}}{n_{\mathsf{B}}!}\,\frac{\bar{x}_{\mathsf{C}}^{n_{\mathsf{C}}}}{n_{\mathsf{C}}!}\;,\;\mathbf{n}=(n_{\mathsf{A}},n_{\mathsf{B}},n_{\mathsf{C}})\in\mathbb{N}^3\,,\;\text{and}\\ \overline{P}_{\mathbf{n}} \;&=\; N\,\overline{p}_{\mathbf{n}}\;\;\text{with}\;\;N\;=\;\left(\sum_{i=0}^{\min\{a_0,b_0\}}\overline{p}_i\right)^{-1}\;,\end{aligned} \tag{4.132}$$

where N is the normalization factor. The equilibrium concentrations, $\bar{x}_{\mathsf{A}}=\bar{x}$, $\bar{x}_{\mathsf{B}}=\vartheta_0+\bar{x}$, and $\bar{x}_{\mathsf{C}}=a_0-\bar{x}$, are readily obtained from the relation

$$K^{-1} \;=\; \frac{[C]}{[A]\cdot[B]} \;=\; \frac{\bar{x}_{\mathsf{C}}}{\bar{x}_{\mathsf{A}}\cdot\bar{x}_{\mathsf{B}}} \;=\; K_b\;,$$

which for $n_{\mathsf{A}}(0)=a_0$ and $n_{\mathsf{B}}(0)=b_0$ with $b_0\geq a_0$, and $n_{\mathsf{C}}(0)=0$, yields

$$\bar{x} \;=\; \frac{1}{2}\Big(a_0-b_0-K+\sqrt{(a_0+b_0+K)^2-4a_0b_0}\Big) \tag{4.133}$$

for the equilibrium concentration of $\mathbf{A}$. In order to generalize to arbitrary a_0 and b_0 values we need only replace $\vartheta_0\Leftrightarrow|b_0-a_0|$ and $a_0\Leftrightarrow\min\{a_0,b_0\}$.

In earlier work on bimolecular chemical reactions expectation value and variance of $\mathcal{X}_{\mathsf{A}}$ was derived by means of probability generating functions [84]:

$$\begin{aligned}\mathrm{E}(\overline{\mathcal{X}}_{\mathsf{A}}) \;&=\; \mu_{\bar{a}} \;=\; K\,\frac{a_0}{\vartheta_0+1}\cdot\frac{{}_1F_1(-a_0+1;\vartheta_0+2;-K)}{{}_1F_1(-a_0;\vartheta_0+1;-K)}\;\;\text{and}\\ \mathrm{var}(\overline{\mathcal{X}}_{\mathsf{A}}) \;&=\; -\mu_{\bar{a}}^2\;-\;(\vartheta_0+K)\,\mu_{\bar{a}}\;+\;a_0\,K\;.\end{aligned} \tag{4.134}$$

In figure 4.25 we consider the stochastic equilibrium in form of the one standard deviation band around the expectation value, $\mathrm{E}(\overline{\mathcal{X}}_{\mathsf{A}})\pm\sigma(\overline{\mathcal{X}}_{\mathsf{A}})$. As expected the relative width of this band becomes smaller with increasing numbers of molecules in the sense of an approximate $\sqrt{N}$-law. The dependence of the probability density on the dissociation constant K for fixed values a_0 and b_0 yields a monotonous increase of the expectation value $\mathrm{E}(\overline{\mathcal{X}}_{\mathsf{A}})$ from $\lim_{K\to0}\mathrm{E}(\overline{\mathcal{X}}_{\mathsf{A}})=0$ to $\lim_{K\to\infty}\mathrm{E}(\overline{\mathcal{X}}_{\mathsf{A}})=a_0$. In contrast to the first order sys-

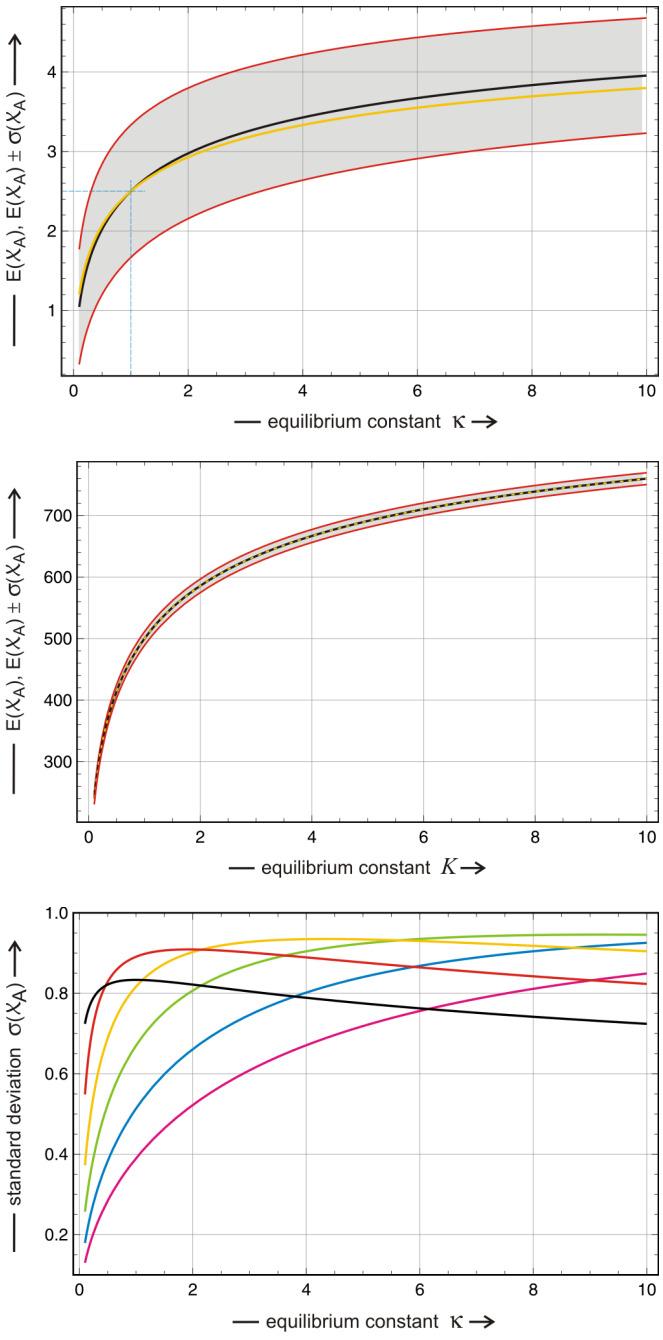


Fig. 4.26 Continued on next page.

Fig. 4.26 Equilibrium of the bimolecular conversion reaction. The first two plots show the equilibrium expectation values $\mathrm{E}\big(\overline{\mathcal{X}}_{\mathsf{A}}(K)\big)$ (black) embedded in the one standard deviation zone $\mathrm{E}(\overline{\mathcal{X}}_{\mathsf{A}}) \pm \sigma(\overline{\mathcal{X}}_{\mathsf{A}})$ (gray with red borders), in addition the deterministic solution $\bar{a}(K)$ (yellow) is shown. Choice of parameters: $a_0 = 5, b_0 = 5$ (upper plot) and $a_0 = 1000, b_0 = 1000$ (middle plot). In the upper plot we see that the curves for $\mathrm{E}\big(\overline{\mathcal{X}}_{\mathsf{A}}(K)\big)$ and $\bar{a}(K)$ cross at the point $K = 1$ as indicated by the blue dotted lines. The third plot shows the standard deviation as a function of the equilibrium constants, $\sigma\big(\overline{\mathcal{X}}_{\mathsf{A}}(K)\big)$. These curves start from $\sigma = 0$ for $K = 0$, pass a maximum and approach zero in the limit $K \to \infty$. Choice of parameters: $a_0 = 5$, $b_0 = 5$ (black), 10 (red), 20 (yellow), 40 (green), 80 (blue), and 150 (magenta).

tem $\mathbf{A} \rightleftharpoons \mathbf{B}$ the expectation value $\mathrm{E}(\overline{\mathcal{X}}_{\mathsf{A}})$ does not coincide with deterministic solution:

$$\bar{a}(a_0, b_0, K) \;=\; \frac{1}{2}\Big(a_0 - b_0 - K + \sqrt{(a_0 - b_0 - K)^2 + 4a_0K}\Big)\;. \qquad (4.133')$$

There is a small but recognizable difference between $\mathrm{E}(\overline{\mathcal{X}}_{\mathsf{A}}; K)$ and $\bar{a}(K)$ for $a_0 = b_0 = 5$, which becomes very small already at moderate particle numbers $(a_0, b_0) > 10$ where the two curves coincide within the line width. The limit of large b_0-values is known as pseudo first order condition with

$$\bar{a} \;\approx\; \tilde{a}(a_0, b_0, K) \;=\; a_0\,\frac{K}{b_0 + K} \quad \text{for} \quad b_0 \gg a_0\;. \qquad (4.135)$$

A factor of $b_0 = 100 \cdot a_0$ is sufficient to make all three curves, $\mathrm{E}(\overline{\mathcal{X}}_{\mathsf{A}}; K)$, $\bar{a}(K)$, and $\tilde{a}(K)$ practically indistinguishable.

Variance and standard deviation of $\mathrm{E}(\overline{\mathcal{X}}_{\mathsf{A}}; K)$ adopt the value zero at both limits, $\lim_{K\to 0} \mathrm{var}(\overline{\mathcal{X}}_{\mathsf{A}}) = 0$ and $\lim_{K\to\infty} \mathrm{var}(\overline{\mathcal{X}}_{\mathsf{A}}) = 0$, and pass a maximum at some intermediate value of K. For constant a_0 the height of the maximum and the position along the K-axis increase with increasing values of b_0. We remark that the equilibrium constant K for the reaction $\mathbf{C} \rightleftharpoons \mathbf{A} + \mathbf{B}$ is not dimensionless, $[K] = [\text{mole} \times l^{-1}] = [\text{numbers of particles}]$, as it was in the first order scenario $\mathbf{A} \rightleftharpoons \mathbf{B}$, an hence analogous scenarios are expected to be observed for equilibrium constants that are scaled with particle numbers.

Reversible bimolecular conversion reaction: $\mathbf{A} + \mathbf{B} \rightleftharpoons \mathbf{C} + \mathbf{D}$. For the purpose of comparison we consider here also the reversible bimolecular conversion reaction (4.1i). As shown in section 4.2.1 three conservation relations reduce the four random variables counting the molecules of class $\mathbf{A}, \mathbf{B}, \mathbf{C}$, and $\mathbf{D}$ to a single one. As initial conditions we choose $\mathcal{X}_{\mathsf{A}}(0) = a_0$, $\mathcal{X}_{\mathsf{B}}(0) = b_0$, $\mathcal{X}_{\mathsf{C}}(0) = c_0$, and $\mathcal{X}_{\mathsf{D}}(0) = d_0$, we make the assumptions $c_0 = 0$, $d_0 = 0$, and introduce $b_0 - a_0 = \vartheta_0$ for the calculations reported here:
$\mathcal{X}_{\mathsf{A}}(t) = n(t)\,,\ \mathcal{X}_{\mathsf{B}}(t) = \vartheta_0 + n(t)\,,\ \mathcal{X}_{\mathsf{C}}(t) = a_0 - n(t)\,,$ and $\mathcal{X}_{\mathsf{D}}(t) = a_0 - n(t)\;.$

The equilibrium probability distribution for this system was calculated by means of the probability generating function [84] and the results for expectation value and variance are:

$$\begin{aligned} \mathrm{E}(\overline{\mathcal{X}}_{\mathsf{A}}) &= \mu_{\bar{a}} = K\,\frac{a_0^2}{\vartheta_0+1}\cdot\frac{{}_2F_1(-a_0+1;-a_0+1;\vartheta_0+2;K)}{{}_2F_1(-a_0;-a_0;\vartheta_0+1;K)}\,, \\ \mathrm{var}(\overline{\mathcal{X}}_{\mathsf{A}}) &= \begin{cases} \mu_{\bar{a}}^2\cdot\frac{b_0^2}{a_0+b_0-1} & \text{if } K=1\,, \\ -\mu_{\bar{a}}^2 - \frac{\vartheta_0+2a_0K}{1-K}\,\mu_{\bar{a}} + \frac{a_0^2K}{1-K} & \text{if } K\neq 1\,. \end{cases} \end{aligned} \tag{4.136}$$

For the purpose of comparison we calculate the deterministic value too:

$$\bar{a}(a_0,b_0,K) = \frac{1}{2(K-1)}\left(\vartheta_0+2a_0K-\sqrt{\vartheta_0^2+4a_0b_0K}\,\right)\,, \tag{4.137}$$

and for $K_b=1$ the equation simplifies to $\bar{a}=a_0^2/(a_0+b_0)$. The illustrative example in figure 4.26 shows an overall picture that is very similar to the associations reaction with two significant differences: (i) the equilibrium constant is dimensionless and this implies that the same values of K can be used for particle numbers and other units to observe the influence on the various phenomena, and (ii) the reaction system exhibits a kind of symmetry at the equilibrium constant $K=1$, where the expectation value and the deterministic equilibrium concentration adopt the same values: $\mathrm{E}(\overline{\mathcal{X}}_{\mathsf{A}};a_0,b_0,1)=\bar{a}(a_0,b_0,1)$. As in the previous example the difference between the the stochastic and the deterministic values becomes very small at relatively small particle numbers, $(a_0,b_0)>10$, already.

4.3.3.3 Autocatalytic reaction

In this section we are dealing with our last example of a chemical reaction for which analytical solutions are available: the bimolecular or first order autocatalytic reaction presented in (4.1g). Here, we present a solution of the master equation, which makes use of the Laplace transform [20]. As in our previous examples the backtransformation into probability space when done analytically provides strong restrictions to the solvable cases.

The reaction for first order autocatalysis in a closed system

$$\mathbf{A} + \mathbf{X} \underset{l}{\overset{k}{\rightleftharpoons}} 2\,\mathbf{X}\,, \tag{4.138}$$

is described by the chemical master equation

$$\begin{aligned}\frac{\mathrm{d}P_n(t)}{\mathrm{d}t} &= k\,(n+1)\,(n_0-n-1)\,P_{n+1}(t) + \\ &+ l\,(n_0-n+1)\,(n_0-n)\,P_{n-1}(t) - \\ &- \Big(k\,n\,(n_0-n) + l\,(n_0-n)\,(n_0-n-1)\Big)\,P_n(t)\end{aligned} \tag{4.139}$$

with $\mathcal{X}_\mathsf{A}(t) = n(t)$ chosen as the single independent stochastic variable, since $\mathcal{X}_\mathsf{X}(t) = n_0 - n(t)$ or $\mathcal{X}_\mathsf{A}(t) + \mathcal{X}_\mathsf{X}(t) = n_0$. As initial conditions we choose $a_0 = n(0)$ and $P_n(0) = \delta_{n,n(0)}$ where we require $x_0 = n_0 - n(0) = n_0 - a_0 > 0$, because otherwise we obtain $\mathrm{d}P_n(0)/\,\mathrm{d}t = 0\,\forall\, n \in [\,0, n_0 - 1], n \in \mathbb{N}$. No reaction takes place if no autocatalyst is present and then the probability density is constant. This is readily verified in the master equation where $n_0 = n(0) \equiv a_0$ or $x_0 = 0$ yields

$$\frac{\mathrm{d}P_n(t)}{\mathrm{d}t} = k\,(n+1)\,(n_0-n-1)\,P_{n+1}(t)\ ,$$

and $P_{n+1}(0) \neq 0$ if and only if $n = a_0 - 1$ where $P_{a_0-1+1}(0) = 1$ but then $n_0 - a_0 + 1 - 1 = 0$. Of course, the same result follows from the deterministic equation: $x_0 = x(0) = 0$ implies no reaction. Another consequence of the autocatalytic process is the fact that the last molecule **X** cannot be converted into an **A** molecule, because two **X** molecules are required for the reaction, and this is reflected by the definition of the probability density.

The master equation can be solved through Laplace transformation in full analogy to the procedure described for the association reaction [20]. The transformation is facilitated by a change in the time axis, $t \Rightarrow t/k$, by the introduction of a dimensionless equilibrium constant, $K = l/k$, as well as step-up and step down transition probabilities according to (3.96):[52]

$$w_n^+ = K\,(n_0-n)\,(n_0-n-1) \quad \text{and} \quad w_n^- = n\,(n_0-n)\ . \tag{4.140}$$

and the master equation takes on the same form as shown previously

$$\frac{\mathrm{d}\mathbf{P}(t)}{\mathrm{d}t} = \mathrm{W}\cdot\mathbf{P}(t)\ , \tag{4.123'}$$

and the matrix W is identical with the only difference that the state $n = n_0$ does not exist or – better expressed – has probability zero, $P_{n_0} = 0$, because the state with $\mathcal{X}_\mathsf{X} = 0$ is an absorbing barrier that, however, cannot be reached from the state $\mathcal{X}_\mathsf{X} = 1$ for the reason mentioned above.

With the usual definition $P_n(t) = 0\,\forall\, n \notin [0, n_0 - 1]$ equation (4.123') represents a linear system of n_0 equations that may be solved by applying a Laplace transform to the components

[52] The difference in the step down transition probabilities is a result of the different stoichiometry of the association reaction and the autocatalytic reaction as discussed already in section 4.1.1 and figure 4.1.

$$V_n(s) \;=\; \int_0^\infty P_n(t)\, e^{-st}\, \mathrm{dt}\ . \tag{4.124'}$$

The same procedure as described in the previous section yields the solution

$$\mathbf{V} \;=\; \Big(s\,\mathbb{I} \,-\, \mathrm{W}\Big)^{-1} \cdot \mathbf{P}_0\ . \tag{4.126'}$$

The initial condition $P_n(0) = \delta_{n,a_0}$ simplifies the calculation of the transformed probability density $V_n(s)$ and allows for the derivation of a closed solution:[53]

$$V_n(s) \;=\; \frac{(-1)^{n+a_0}\, \mathrm{M}_{a_0+1,n+1}(s)}{\det(s\mathbb{I} - \mathrm{W})}\ , \tag{4.141}$$

where $\mathrm{M}_{a_0+1,n+1}$ is a minor of the matrix $(s\mathbb{I} - \mathrm{W})$.

For the irreversible reaction $\mathbf{A} + \mathbf{X} \to 2\mathbf{X}$ – the case $l = K = 0$ – the reverse reaction is precluded and the vectors $\mathbf{V}$ and $\mathbf{P}_0$ as well as the matrix W have the dimensions $(a_0 + 1) \times 1$ and $(a_0 + 1) \times (a_0 + 1)$, respectively. The inverse Laplace transformation is performed in analogy to the previous section. A relevant difference, however, occurs because, of the degeneracy of some eigenvalues of matrix W.

Since all step-up transitions are forbidden $w_n^+ = 0\,\forall\, n \in [0, a_0]$ the matrix W is upper-diagonal and has the eigenvalues

$$\lambda_j \;=\; -\,w_j^-\ ;\ j = 0, 1, \ldots, a_0 \quad \text{with}\ \ \lambda_j = \lambda_{a_0-j}\ \ \text{since}\ \ w_j^- = w_{a_0-j}^-\ .$$

Two cases are distinguished: (1) for $x_0 \geq a_0$ the eigenvalues of W are distinct and (2) degenerate pairs may occur when $x_0 < a_0$, in particular all eigenvalues $\lambda_j, \lambda_{n_0-j}$ are degenerate for $j \in [x_0, a_0]$ except $\lambda_{n_0/2}$ if n_0 is even.

If all eigenvalues of W are distinct, the probability distribution can be obtained from the Laplace transform by means of the Heaviside expansion theorem [82]:

$$P_n(t) \;=\; \sum_{j=0}^{n_0-1} \lim_{s\to\lambda_j} (s - \lambda_j)\, e^{st}\, V_n(s)\ .$$

The sum covers all eigenvalues of W. The derivation of the expansion theorem requires non-degeneracy of the roots $\lambda_j = s_j$ in the general formulation

$$f(t) \;=\; \mathcal{L}^{-1}\left(\frac{P(s)}{Q(s)}\right) \;=\; \sum_{k=1}^{n} \frac{P(s_k)}{Q'(s_k)}\, e^{s_k t} \quad \text{with}\ \ Q'(s_k) \;=\; \left.\frac{\mathrm{d}Q(s)}{\mathrm{ds}}\right|_{s=s_k}\ ,$$

where $P(s)$ and $Q(s)$ are polynomials in s of degree m and n respectively, which fulfil $n > m$. Then follows for every simple root s_k of $Q(s)$:

---

[53] In general, the calculation of determinants and minors is highly nontrivial as is the subsequent inversion of the Laplace transform but thanks to the sharp initial conditions applied here all steps can be performed analytically.

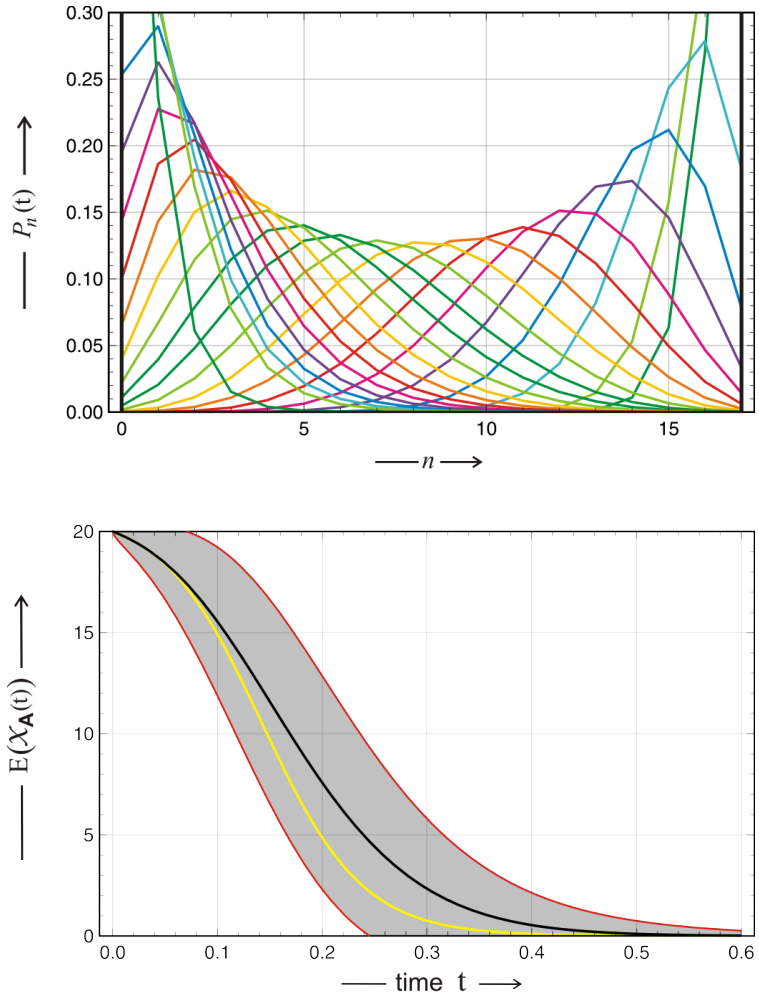


Fig. 4.27 Irreversible bimolecular autocatalytic reaction $\mathbf{A} + \mathbf{X} \to 2\mathbf{X}$. The plot shows the probability distribution $P_n(t) = P\big(\mathcal{X}_\mathsf{A}(t) = n\big)$ describing the number of molecules of species **A** as a function of time and calculated by the equations in table 4.1. Parameter choice: $k = 1\,[\mathrm{N}^{-1}{\cdot}\mathrm{V}{\cdot}\mathrm{sec}^{-1}]$, $a_0 = 17$, $x_0 = 5$, and sampling times: $t = 0$ (black), 0.005 (chartreuse), 0.01 (green), 0.02 (turquoise), 0.03 (blue), 0.04 (violet), 0.06 (purple), 0.08 (magenta), and ∞ (red). In the lower plot we show the expectation value $\mathrm{E}\big(\mathcal{X}_\mathsf{A}(t)\big)$ (black) together with the band $\mathrm{E} \pm \sigma$ (red) and the deterministic expectation value (yellow). The areas with where the calculated values are probabilistically meaningless, $\mathrm{E} + \sigma > a_0$ and $E - \sigma < 0$, are clipped. Parameter choice: $a_0 = 20$, $x_0 = 1$, and $k = 1\,[\mathrm{N}^{-1}{\cdot}\mathrm{V}{\cdot}\mathrm{sec}^{-1}]$.

$$\mathrm{Res}\Big(F(s)\,e^{st}\,,\,s_k\Big) \;=\; \lim_{s\to s_k}\frac{s-s_k}{Q(s)-Q(s_k)}\,P(s)\,e^{st} \;=\; \frac{P(s_k)}{Q'(s_k)}\,e^{s_k t}\;.$$

Extension to the entire domain of s is done by summation.

Degenerate eigenvalues can be handled by singular perturbation theory. For this goal we replace w_n^- by

$$\eta_j \;=\; j\,(n_0-j+\epsilon)\;,$$

and after some calculation one finds for the solution in the Laplace space

$$V_n(s|\epsilon) \;=\; \frac{(-1)^{n+a_0}\prod_{j=n+1}^{a_0}(-\eta_i)}{\prod_{i=n}^{a_0}(s+\eta_i)}\;. \qquad (4.141')$$

Inverse Laplace transformation and evaluation of the products yields the result that still depends on ϵ:

$$\begin{aligned} P_n(t|\epsilon) \;&=\; \sum_{j=n}^{a_0}\frac{(-1)^{n+a_0}\prod_{i=n+1}^{a_0}(-\eta_i)}{\prod_{i=n}^{j-1}\prod_{i=j+1}^{a_0}}\,e^{-\eta_j t} \;= \\ &=\; \frac{(x_0+\epsilon)_{a_0-n}}{n!}\sum_{j=n}^{a_0}A_{j,n}^{(x_0,a_0)}(\epsilon)\,e^{-j(n_0-j+\epsilon)t} \quad\text{with} \qquad (4.142) \\ A_{j,n}^{(x_0,a_0)}(\epsilon) \;&=\; (-1)^{a_0+j}\binom{a_0}{j}\frac{(j-n+1)_n\,(2j-n_0-\epsilon)}{(j+n-n_0-\epsilon)_{a_0-n+1}}\;. \end{aligned}$$

and $(x)_n = \Gamma(x+n)/\Gamma(x)$ is the rising Pochhammer polynomial.

The next and final step is the evaluation of the limit $\epsilon\to 0$ where three different special cases have to be distinguished. For this purpose we define two auxiliary functions $B_{j,n}^{(x_0,a_0)}$ and $C_{j,n}^{(x_0,a_0)}(t)$:

$$B_{j,n}^{(x_0,a_0)} \;=\; \lim_{\epsilon\to 0}\frac{(x_0+\epsilon)_{a_0-n}}{n!}\,A_{j,n}^{(x_0,a_0)}(\epsilon)\;,$$

The function $B_{j,n}^{(x_0,a_0)}$ is independent of time and appears as coefficient of an exponential decrease term in the final expression for $P_n(t)$. For convenience to cases are distinguished:

$$B_{j,n}^{(x_0,a_0)} \;=\; \begin{cases} \frac{(-1)^{j+n}a_0!(n_0-n-1)!(x_0-j-1)!(n_0-2j)}{n!(x_0-1)!(a_0-j)!(j-n)!(n_0-j-n)!} & \text{if } j+n\le n_0\,, \\ \frac{(-1)^{a_0+j}a_0!(j+n-n_0-1)!(x_0-j-1)!(2j-n_0)}{n!(x_0-1)!(a_0-j)!(j-n)!(j-x_0)!} & \text{if } j+n>n_0\,. \end{cases} \qquad (4.143)$$

The second function $C_{j,n}^{(x_0,a_0)}(t)$ is more involved and results from accounting for the degeneracies in the eigenvalues:

$$\begin{aligned} C_{j,n}^{(x_0,a_0)}(t) \;&=\; (-1)^{n-x_0}\,\frac{(n_0-n-1)!a_0!}{n!(x_0-1)!(a_0-j)!(j-n)!(j-x_0)!(n_0-j-n)!}\,\cdot \\ &\cdot\,\Big((n_0-2j)^2t\,+\,2\,-\,(n_0-2j)(H_{j-n}-H_{a_0-j}-H_{n_0-j-n}+H_{j-x_0})\Big)\,. \end{aligned} \qquad (4.144)$$

Table 4.1 The probability density of the first order irreversible autocatalytic reaction $\mathbf{A}+\mathbf{X} \to 2\mathbf{X}$. For $x_0 > a_0$ all eigenvalues of matrix W are distinct and the probability density is obtained by a simple sum of the contributions of individual exponential decay modes. The expressions are taken from [20].

Case	Range	Probability density $P_n(t)$
$x_0 > a_0$	$[0, a_0]$	$\sum_{j=n}^{a_0} B_{j,n}^{(x_0,a_0)}\, e^{-k\,j(n_0-j)t}$
$x_0 \le a_0$	$[0, x_0[$	$\sum_{j=n}^{x_0-1} B_{j,n}^{(x_0,a_0)}\, e^{-k\,j(n_0-j)t} + \sum_{j=x_0}^{\lfloor n \rfloor} \frac{C_{j,n}^{(x_0,a_0)}(t)\, e^{-k\,j(n_0-j)t}}{1+\delta_{j,n_0-j}}$
	$[x_0, \lfloor n \rfloor]$	$\sum_{j=n}^{\lfloor n \rfloor} \frac{C_{j,n}^{(x_0,a_0)}(t)\, e^{-k\,j(n_0-j)t}}{1+\delta_{j,n_0-j}} + \sum_{j=n}^{n_0-n+1} B_{j,n}^{(x_0,a_0)}\, e^{-k\,j(n_0-j)t}$
	$]\lfloor n \rfloor, a_0]$	$\sum_{j=n_0-n+1}^{a_0} B_{j,n}^{(x_0,a_0)}\, e^{-k\,j(n_0-j)t}$

By H_n we denote here the harmonic numbers: $H_n = \sum_{k=1}^{n} \frac{1}{k}$. The derivation of the final expressions is technically quite involved since the case with degeneracies of eigenvalues has to be split in subcases requiring different handling of summations and we refer to the literature for details [20]. In table 4.1 we present the probability densities for the different cases and subcases. There we have already transformed back to the original time $t \Rightarrow k\,t$. An example of the time dependent probability density $P_n(t)$ of the irreversible first order autocatalytic reaction is shown in figure 4.27.

The expectation value of the number of molecules $\mathbf{A}$ can be computed also by means of auxiliary functions with some labor [20]. Direct calculation of mean and variance, however, is less sophisticated:

$$\begin{aligned} \mathrm{E}\Big(\mathcal{X}_{\mathsf{A}}(t)\Big) &= \sum_{n=0}^{a_0} n\, P_n(t) \quad \text{and} \\ \mathrm{var}\Big(\mathcal{X}_{\mathsf{A}}(t)\Big) &= \sum_{n=0}^{a_0} n^2\, P_n(t) \;-\; \mathrm{E}\Big(\mathcal{X}_{\mathsf{A}}(t)\Big)^2 \,. \end{aligned}$$

An illustrative example is shown in figure 4.27 where we can see also a substantial difference between the deterministic solution and the expectation value.

In principle, the master equation of the reversible first-order autocatalytic reaction (4.123) could be handled by the same procedure as the irreversible reaction. In the irreversible case the eigenvalues of matrix W are available in analytical form. Since this does not seem to be possible for the reversible case, little is gained by the Laplace transform. We discuss $\mathbf{A} + \mathbf{X} \rightleftharpoons 2\mathbf{X}$ as an example for numerical simulations in section 4.6.4. There we discuss in detail also the differences between the stochastic and the deterministic solutions. The stationary solution of the reversible reaction is, nevertheless,

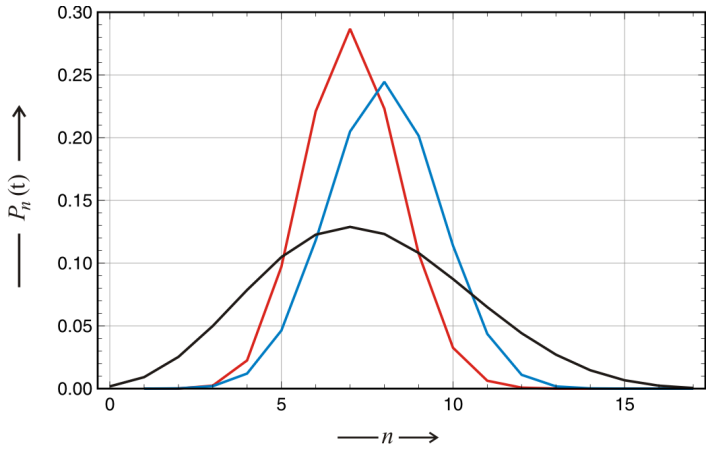


Fig. 4.28 Probability densities of bimolecular reactions. Compared are the widths of the densities $P\big(\mathcal{X}_{\mathsf{A}}(t_m)\big)$ in the middle of the state space that were calculated with equations Parameter choice: $(t_m = 0.09, k = 1)$ for $\mathbf{A} + \mathbf{X} \to 2\mathbf{X}$ (black), $(t_m = 0.4375, k = 0.1)$ for $\mathbf{A} + \mathbf{B} \to 2\mathbf{C}$ (blue), and $(t_m = 0.525, k = 0.1)$ for $2\,\mathbf{A} \to 2\mathbf{C}$ (red).

readily computed by applying the results for the longtime limit:

$$\mathbf{0} \;=\; \mathrm{W} \cdot \bar{\mathbf{P}} \quad \text{with} \quad n \in [0, n_0[\text{ and } K > 0 \;, \tag{4.123'}$$

which can be done for any stationary univariate master equation by means of (3.100). By inserting the expressions from (4.140) we find

$$\bar{P}_n^{(auto)} \;=\; \binom{n_0}{n} \frac{K^n}{(1+K)^{n_0} - K^{n_0}} \tag{4.145}$$

This result is to be compared with the equilibrium of the monomolecular reaction $\mathbf{A} \rightleftharpoons \mathbf{X}$ that was calculated in section 4.3.2.2

$$\bar{P}_n^{(mono)} \;=\; \binom{n_0}{n} \frac{K^n}{(1+K)^{n_0}} \;.$$

We recognize the difference of the density distributions for the two reactions at equilibrium is the fact that a single $\mathbf{X}$ molecule cannot be converted into an $\mathbf{A}$ molecule and a different normalization is required. This deviation disappears with increasing values of n_0 as fast as $n(n-1)$ approaches n^2.

In figure 4.28 we compare the width of the probability densities of all three irreversible bimolecular reactions studied here: $\mathbf{A} + \mathbf{X} \to 2\mathbf{X}$, $\mathbf{A} + \mathbf{B} \to \mathbf{C}$, and $2\,\mathbf{A} \to \mathbf{C}$. All three reactions start from a sharp distribution, $P_n(0) = \delta_{n,a0}$, and progress to a sharp distribution, $\lim_{t\to\infty} P_n(t) = \delta_{n,0}$. In order to

make the densities comparable, we consider them in the middle of the state space, $\mathrm{E}\big(\mathcal{X}_{\mathsf{A}}(t)\big) \approx a_0/2$. The autocatalytic process is characterized by two differences in comparison with the other two reactions: (i) the distribution is much broader and (ii) the time at which the distribution passes the middle of the state space is much shorter. Both findings are a result of the self-enhancement in autocatalysis. Fluctuations are larger and the rate of the reaction is accelerated.

4.3.4 Stochastic enzyme kinetics

Michaelis-Menten kinetics is more complex than the the examples treated here so far, since even the simple mechanism, $\mathbf{S} + \mathbf{E} \rightleftharpoons \mathbf{C} \rightarrow \mathbf{E} + \mathbf{P}$ with $\mathbf{C}$ denoting the enzyme-substrate complex, $\mathbf{C} \equiv \mathbf{S}{\cdot}\mathbf{E}$, cannot be reduced without approximation to a problem with a single independent variable. Instead, we have to deal with two random variables, for example $\mathcal{X}_{\mathsf{S}}$ and $\mathcal{X}_{\mathsf{E}}$, counting substrate and enzyme molecules, respectively, and with a bivariate probability density: $P_{e,n}(t)$ where e denotes the number of free enzyme molecules $\mathbf{E}$ and n represents the number of substrate molecules $\mathbf{S}$. The analytical model we introduce here is taken from the literature [13]. It is based on the assumption that only a single enzyme molecule – free or bound in the complex, E or C, respectively[54] – is present, and this is interpreted as the consideration of a sufficiently small volume such that it contains one or no enzyme molecule. Present day spectroscopic techniques made it possible to observe and study single enzyme molecules (section 4.4.1) and the model presented here found a physical realization in experimental setups with single enzyme molecules that are immobilized in compartments or on membranes.

The basic steps of irreversible substrate to product conversion, $\mathbf{S} \rightarrow \mathbf{P}$, are:

$$\ldots \overset{k_2}{\rightarrow} n\,\mathsf{S} + \mathsf{E} \underset{l_1}{\overset{k_1 \cdot n}{\rightleftharpoons}} (n-1)\,\mathsf{S} + \mathsf{C} \overset{k_2}{\rightarrow} (n-1)\,\mathsf{S} + \mathsf{E} + \mathsf{P} \underset{l_1}{\overset{k_1 \cdot (n-1)}{\rightleftharpoons}} \ldots ,$$

where n is not the stoichiometric coefficient but the number of substrate molecules that are ready for conversion. In figure 4.29 we show the entire state space for a single enzyme molecule, $e \in \{0, 1\}$ and $n \in [0, n_0], n \in \mathbb{N}$. It is straightforward to write down a master equation for this scheme:

$$\begin{aligned} \frac{\mathrm{d}P_{e,n}(t)}{\mathrm{dt}} \;=\;& l_1(2-e)P_{e-1,n-1}(t) \;+\; k_2(2-e)P_{e-1,n}(t) + \\ &+\; k_1(e+1)(n+1)P_{e+1,n+1}(t) - \\ &-\; \big(k_1\, e\, n + (l_1 + k_2)(1-e)\big)P_{e,n}(t) \;, \end{aligned} \tag{4.146}$$

[54] Since we mean here molecules rather than molecular species we use normal instead of bold fonts. The diagram sketches the state space and is called a *reaction scheme* in order to point at the difference to a reactions mechanism.

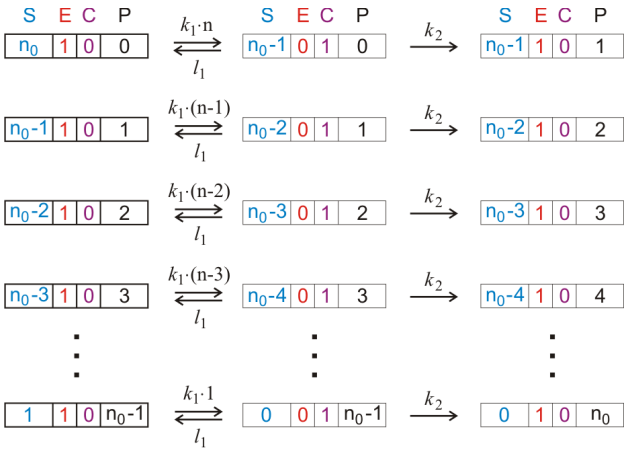


Fig. 4.29 Scheme of the Michaelis-Menten mechanism with a single enzyme molecule. We show the irreversible conversion of n substrate molecules into n product molecules that occurs in $2n$ individual reaction steps. The boxes contain the numbers of molecules of the four species: substrate **S** (blue), enzyme **E** (red), enzyme-substrate complex **C** $\equiv$ **S**·**E** (purple), and product **P** (black). All states in the third column are identical with the states of the first column in the next row except the initial and the final state and hence the reaction scheme consists of a single line.

with the initial conditions $P_{e,n}(0) = \delta_{e,1} \cdot \delta_{n,n_0}$.[55] Since the conversion steps are irreversible the final state is determined by the limiting density

$$\lim_{t\to\infty} P_{e,n}(t) \;=\; \delta_{e,1} \cdot \delta_{n,0} \;,$$

all substrate molecules S are converted into product P and the enzyme is in the free state E.

A solution of the master equation (4.146) can be derived by means of the marginal probability generating functions:[56]

$$g_e(s,t) \;=\; \sum_{n=0}^{n_0+e-1} s^n \, P_{e,n}(t) \quad \text{with} \quad e \in \{0,1\},\, t \geq 0 \;. \tag{4.147}$$

Equations (4.146) and (4.147) are converted into a system of partial differential equations:

[55] Here and in the following paragraphs all probability densities and generating functions with index values outside the domains, $e \notin \{0,1\}$ and $n \notin [0,n]$ are zero.

[56] This approach is meaningful in case one of the two random variables is restricted to very few values, here $\mathcal{X}_{\mathsf{E}} = \{0,1\}$. The use of marginal densities avoids the occurrence of second order partial derivatives that create the difficulties encountered in solving the master equations of second order reactions.

$$\begin{aligned}\frac{\partial g_e(s,t)}{\partial t} \;&=\; k_1(e+1)\frac{\partial g_{e+1}(s,t)}{\partial s} \;-\; k_1\, e\, s\frac{\partial g_e(s,t)}{\partial s} - \\ &-\; (l_1+k_2)(1-e)\, g_e(s,t) + \\ &+\; (l_1+k_2)(2-e)s\, g_{e-1}(s,t) \quad \text{with} \quad e\in\{0,1\}\;.\end{aligned} \tag{4.148}$$

The solution of the PDE (4.148) is obtained in terms of a pair of generating functions:

$$\begin{aligned}g_0(s,t) \;&=\; \gamma_1\, e^{-l_1(s-1)/k_2}\cdot e^{-k_2 t} \;+\; \gamma_2\frac{l_1+k_2}{l_1 s+k_2}\, e^{-(k_1+k_2)t} + \\ &+\; \sum_{i=1}^{2}\sum_{n=0}^{\infty}\chi_i^{(n)}\left(\frac{k_2-(k_2+\lambda_i^{(n)})\, s}{-\lambda_i^{(n)}}\right)^{q_i^{(n)}+1} e^{\lambda_i^{(n)}t}\;,\end{aligned} \tag{4.149a}$$

$$\begin{aligned}g_1(s,t) \;&=\; \gamma_0 \;-\; \gamma_1\, e^{-l_1(s-1)/k_2}\cdot e^{-k_2 t} \;-\; \gamma_2\frac{l_1+k_2}{l_1 s+k_2}\, e^{-(k_1+k_2)t} - \\ &-\; \sum_{i=1}^{2}\sum_{n=0}^{\infty}\chi_i^{(n)}\left(\frac{k_2-(k_2+\lambda_i^{(n)})\, s}{-\lambda_i^{(n)}}\right)^{q_i^{(n)}+1} e^{\lambda_i^{(n)}t}\;,\end{aligned} \tag{4.149b}$$

where the various coefficients γ_i and $\chi_i^{(k)}$ are to be determined from the conditions:

$$g_0(1,t)+g_1(1,t) \;=\; 1\,,\;\; g_0(s,0)=0\,,\quad \text{and}\quad g_1(s,0) \;=\; s^{n_0}\;.$$

The eigenvalues of the $(2n_0+1)\times(2n_0+1)$ transition probability matrix corresponding to the master equation (4.146) are obtained from n_0 quadratic equations:

$$\lambda^2 \;+\; \big(l_1+k_1(n+1)+k_2\big)\,\lambda \;+\; k_1k_2\,(n+1) \;=\; 0 \;\;\forall\;\; n=0,1,\ldots\, n_0-1\,,$$

including a trivial eigenvalue $\lambda=0$:

$$\lambda_{1,2}^{(n)} \;=\; -\frac{1}{2}\left(\big(k_1(n+1)+l_1+k_2\big)\mp\sqrt{\big(k_1(n+1)+l_1+k_2\big)^2-4k_1k_2(n+1)}\right)$$

The exponents $q_i^{(n)}$ are readily obtained from the eigenvalues:

$$q_i^{(n)} \;=\; -\,\frac{\big(\lambda_i^{(n)}\big)^2+(l_1+k_1+k_2)\lambda_i^{(n)}+k_1k_2}{k_1\,\big(k_2+\lambda_i^{(n)}\big)}\;.$$

Although the calculations can be quite involved in practice, the coefficients γ_1 and γ_2 vanish if $l_1\neq 0$, the summations over n contain finite numbers of terms, the $q_i^{(n)}$-values are in the range $[0\le q_i^{(n)}\le n_0-1]$, and the eigenvalues $\lambda_i^{(n)}$ are distinct real numbers.

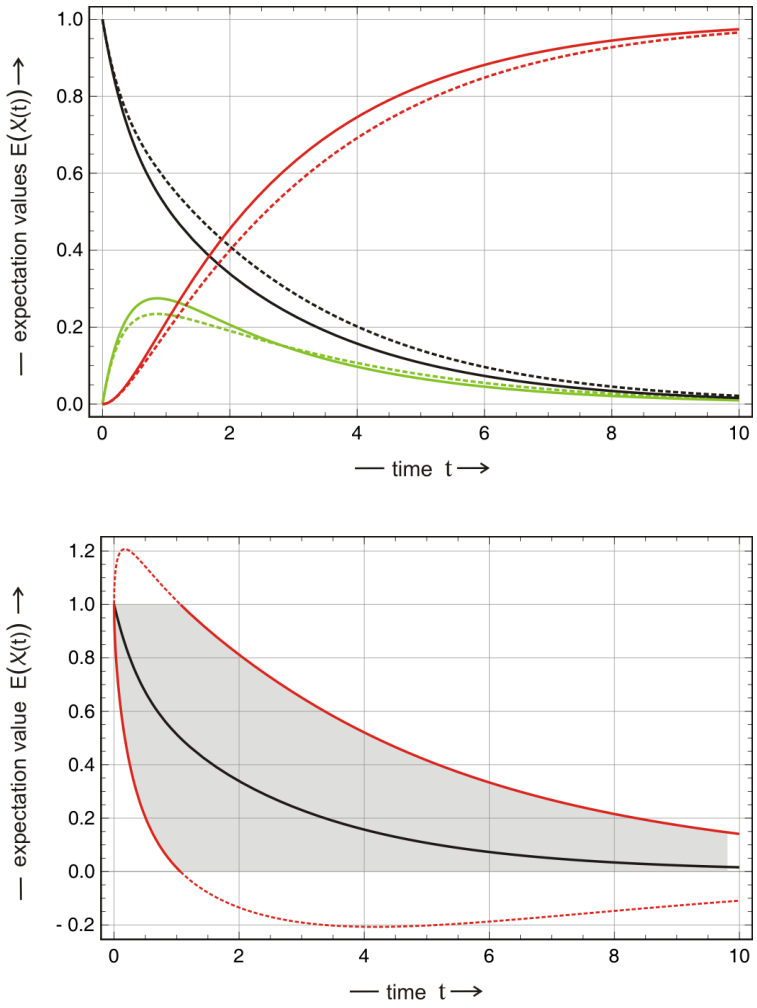


Fig. 4.30 The Michaelis-Menten reaction with single molecules. The full curves in the upper plot shows the expectation values for observing the substrate molecule $\mathrm{E}\big(\mathcal{X}_{\mathsf{S}}(t)\big)$ (black), the substrate-enzyme complex $\mathrm{E}\big(\mathcal{X}_{\mathsf{C}}(t)\big)$ (green), and the product molecule $\mathrm{E}\big(\mathcal{X}_{\mathsf{P}}(t)\big)$ (red), which are compared with the corresponding functions $s(t)$, $c(t)$, and $p\,(t)$ (broken curves) obtained by integration of the kinetic differential equations (section 4.1.2) for the same conditions. The lower plot presents the one standard deviation error band around the expectation value of the substrate concentration: $\mathrm{E}\big(\mathcal{X}_{\mathsf{S}}(t)\big) \pm \sigma\big(\mathcal{X}_{\mathsf{S}}(t)\big)$. The gray hatched zone is the probabilistically meaningful part of the error band.

The probability densities are obtained from the generating functions in the conventional way:

$$P_{e,n}(t) \;=\; \frac{1}{n!}\,\frac{\partial^n g_e(s,t)}{\partial s^n}\,\Big|_{s=0}\,,$$

As an example demonstrate the usage of single molecule probabilities the densities for a single enzyme molecule and n_0 substrate molecules restricted to the conversion of the first substrate molecule are calculated [115]:

$$P_{0,n_0-1}(t) \;=\; \frac{k_1\,n_0}{\lambda_2^{(n_0-1)}-\lambda_1^{(n_0-1)}}\left(e^{\lambda_2^{(n_0-1)}t}-e^{\lambda_1^{(n_0-1)}t}\right) \;\text{ and}$$

$$P_{1,n_0}(t) \;=\; \frac{\lambda_1^{(n_0-1)}+k_1\,n_0}{\lambda_2^{(n_0-1)}-\lambda_1^{(n_0-1)}}\;e^{\lambda_2^{(n_0-1)}t} \;-\; \frac{\lambda_2^{(n_0-1)}+k_1\,n_0}{\lambda_2^{(n_0-1)}-\lambda_1^{(n_0-1)}}\;e^{\lambda_1^{(n_0-1)}t}\,.$$

The calculation of the time of the expected appearance of the first product molecule, ϑ, is a first passage time problem. The probability of recording a product molecule, $P_{\mathsf{P}}(t)$, is simply the probability of having neither a substrate molecule S nor an enzyme-substrate complex C given by the expression $P_{\mathsf{P}}(t) = P\big(\mathcal{X}_{\mathsf{P}}(t)=1\big) = 1 - P_{1,n_0}(t) - P_{0,n_0-1}(t)$. The expectation value of the time ϑ is then obtained from

$$\langle\vartheta\rangle \;=\; \int_0^\infty t\,\frac{\mathrm{d}P_{\mathsf{P}}(t)}{\mathrm{dt}}\,\mathrm{dt} \;=\; -\int_0^\infty t\,\left(\frac{\mathrm{d}P_{1,n_0}(t)}{\mathrm{dt}}+\frac{\mathrm{d}P_{0,n_0-1}(t)}{\mathrm{dt}}\right)\,\mathrm{dt}\,.$$

After some calculations we obtain the final result

$$\langle\vartheta\rangle \;=\; \frac{k_1\,n_0\;+\;l_1\;+\;k_2}{k_1k_2\,n_0} \;=\; \frac{1}{k_1\,n_0}\;+\;\frac{l_1}{k_1k_2\,n_0}\;+\;\frac{1}{k_2}\;, \qquad (4.150)$$

which is easily interpreted: The appearance of the (first) molecule P takes long if (i) the binding rate constant k_1 multiplied by the initial number of substrate molecules is small, or if (ii) the dissociation of the enzyme substrate complex $\mathbf{C} \to \mathbf{S} + \mathbf{E}$ is fast as expressed by a large value of l_1, or if (iii) the rate constant of product formation, k_2, is small.

For the purpose of illustration and comparison with the deterministic model we consider the one enzyme molecule plus one substrate molecule in figure 4.30:

$$P_{00}(t) \;=\; \frac{k_1}{\lambda_2^{(0)}-\lambda_1^{(0)}}\left(e^{\lambda_2^{(0)}t}-e^{\lambda_1^{(0)}t}\right) \;\text{ and}$$

$$P_{11}(t) \;=\; \frac{\lambda_1^{(0)}+k_1}{\lambda_2^{(0)}-\lambda_1^{(0)}}\;e^{\lambda_2^{(0)}t} \;-\; \frac{\lambda_2^{(0)}+k_1}{\lambda_2^{(0)}-\lambda_1^{(0)}}\;e^{\lambda_1^{(0)}t}\,.$$

These probability densities are identical with the expectation values for the different states in the last line of figure 4.29, since the random variables can take on only the values zero and one for $n_0 = 1$:

$$\mathrm{E}\big(\mathcal{X}_{\mathsf{S}}(t)\big) = P_{11}(t),\ \mathrm{E}\big(\mathcal{X}_{\mathsf{C}}(t)\big) = P_{00}(t) \text{ and } \mathrm{E}\big(\mathcal{X}_{\mathsf{P}}(t)\big) = 1 - P_{11}(t) - P_{00}(t)\ ,$$

and in the single molecule case the three variances are readily calculated to be: $\mathrm{var}\big(\mathcal{X}(t)\big) = \mathrm{E}\Big(\mathcal{X}(t) - \mathrm{E}\big(\mathcal{X}(t)\big)\Big)^2$. In figure 4.30 the time dependence of the probabilities is compares with the curves obtained by integration of the kinetic differential equations (section 4.1.2): We find a remarkably good agreement despite the fact that the expectation values refer to single molecule events. The variance, however, is so large that the curves $\mathrm{E}\big(\mathcal{X}(t)\big) \pm \sigma\big(\mathcal{X}(t)\big)$ extend also outside the probabilistic domain: $0 \leq \mathrm{E}\big(\mathcal{X}(t)\big) \leq 1$. Then, it is advisable to restrict the one standard deviation error zone to the meaningful domain.[57]

[57] The same over- and undershooting of the $\mathrm{E}\big(\mathcal{X}(t)\big) \pm \sigma\big(\mathcal{X}(t)\big)$ curves has also been observed in previously discussed cases. For most systems, however, the meaningless parts of the one standard deviation error band are negligibly small.

4.4 Fluctuations and single molecules investigations

The rapid advancements of molecular spectroscopy with respect to signal intensity and temporal resolution within the second half of the twentieth century became the basis for entirely new developments. The old dream to be able to watch single molecules in action and to record single events became true first with the electric current flowing true membranes: The patch-clamp technique made it possible to register opening and closing of single ion channels [188, 345]. Another breakthrough was more general and came from fluorescence spectroscopy: Signals from single molecules were recorded in solution [210, 386] and in the solid state [332, 355] (for reviews see, e.g., [267, 366, 455]), and set the stage the analysis of single molecules. By means of fluorescence the motions of single molecules can now be traced routinely at high spatial and temporal resolution and in highly heterogeneous environments like living cells. A third approach came from applications of scanning tunneling microscopy [42] and opened the possibility for mechanical manipulation of single molecules by means of atomic force microscopy [473]. Particularly illustrative in this context is the mechanochemistry of single nucleic acid molecules [194, 282]. In essence, the single particle approaches may be grouped into three classes: (i) methods to record the states of single molecules in solution, (ii) methods to track the motions of single particles in space, and (iii) methods to manipulate single particles mechanically. The literature on successful single molecule experiments is enormous. Here, we shall focus on two selected issues that require the stochastic methods presented in this monograph, biochemical kinetics of single enzyme molecules, a new method that provides new insights into the mechanism of enzymatic catalysis, and fluorescence correlation spectroscopy, a general technique that allows for recording of single particles.

4.4.1 Single molecule enzymology

The possibility to record signals from single protein molecules and to follow their time dependence is the basis for experimental single molecule enzymology. The insight into the mechanistic details gained by single molecule studies provide answers to a number of questions of the kind: *Are all enzyme molecules in the same conformation and do they react with the same rate parameters or are we dealing with conformational fluctuations, which give rise to dynamical disorder?* or *Are the fluctuations in enzyme turnover and substrate to product conversion in enzymatic reactions larger than in the case of catalysis by small molecules?* In this section we shall present stochastic treatments of the two extended versions of Michaelis-Menten kinetics shown in figure 4.2 and begin with scheme A [378].

Extended Michaelis-Menten mechanism (scheme **A***).* The enzyme molecule can exist in three conformational states, **E**, **S**·**E**, and **E**·**P** (figure 4.2, scheme **A**), and the random variable $\mathcal{X}(t)$ fluctuates between the three states $S_1 = \mathbf{E}$, $S_2 = \mathbf{C} \equiv \mathbf{S}{\cdot}\mathbf{E}$, and $S_3 = \mathbf{D} \equiv \mathbf{E}{\cdot}\mathbf{P}$. The probabilities for the enzyme molecule to be in one of the three states at time t are denoted by

$$\begin{aligned} P_{\mathbf{E}}(t) =& P\big(\mathcal{X}(t) = \mathbf{E}\big), P_{\mathbf{C}}(t) = P\big(\mathcal{X}(t) = \mathbf{C}\big), P_{\mathbf{D}}(t) = P\big(\mathcal{X}(t) = \mathbf{D}\big) \\ & \text{with} \;\; P_{\mathbf{E}}(t) + P_{\mathbf{C}}(t) + P_{\mathbf{D}}(t) \;=\; 1 \;. \end{aligned} \tag{4.151}$$

As in section 4.1.2 we make the assumption of constant substrate and product concentration: $k_1 = k_1'[\mathbf{S}]_0 = k_1' s_0$ and $l_3 = l_3'[\mathbf{P}]_0 = l_3' p_0$. Then, the time dependence of the probabilities is determined by the linear ODE (4.22). Because of the conservation relation it is obtained a superposition of two exponential functions as follows directly from linear algebra:

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{dt}} \;=\; \mathrm{H}\cdot\mathbf{P} \;\;\text{with}\;\; \mathbf{P} = \begin{pmatrix} P_{\mathbf{E}} \\ P_{\mathbf{C}} \\ P_{\mathbf{D}} \end{pmatrix} \;\;\text{and}\;\; \mathrm{H} = \begin{pmatrix} -k_1 - l_3 & l_1 & k_3 \\ k_1 & -k_2 - l_1 & l_2 \\ l_3 & k_2 & -k_3 - l_2 \end{pmatrix}.$$

Solutions of the ODE are obtained in terms of an eigenvalue problem where Λ is a diagonal matrix containing the eigenvalues $\lambda_{1,2}$ and $\lambda_0 = 0$:

$$\begin{aligned} \mathbf{P}(t) \;&=\; \mathrm{B}\cdot\boldsymbol{\varrho}(t) \;\;\text{with}\;\; \mathrm{B}^{-1}\cdot\mathrm{H}\cdot B = \Lambda \;, \\ \frac{\mathrm{d}\boldsymbol{\varrho}}{\mathrm{dt}} \;&=\; \Lambda\cdot\boldsymbol{\varrho} \;\;\text{and}\;\; \boldsymbol{\varrho}(t) \;=\; \exp(\Lambda t)\boldsymbol{\varrho}(0), \end{aligned}$$

where $\boldsymbol{\varrho} = (\varrho_0, \varrho_1, \varrho_2)^{\mathrm{t}}$ contain the right-hand eigenvectors of the matrix H. Backtransformation into the original probabilities eventually yields:

$$\mathbf{P}(t) \;=\; \mathrm{B}\cdot\exp(\Lambda t)\cdot\Big(\mathrm{B}^{-1}\cdot\mathbf{P}(0)\Big) \;.$$

The eigenvalues of matrix H have been calculated already for the deterministic system (4.23). Both eigenvalues have negative real parts and the asymptotically stable stationary state (4.21) corresponds to the macroscopic thermodynamic equilibrium. Here, we are computing normalized probabilities (for the sake of simplicity we use numbers instead of letters as indices, $1 \equiv \mathsf{S}_1 \equiv \mathbf{E}$, $2 \equiv \mathsf{S}_2 \equiv \mathbf{C}$, $3 \equiv \mathsf{S}_3 \equiv \mathbf{D}$, and find [378]:

$$\begin{aligned} \bar{P}_1 \;&=\; \frac{k_2k_3 + k_3l_1 + l_1l_2}{N} \;, \\ \bar{P}_2 \;&=\; \frac{k_3k_1 + k_1l_2 + l_2l_3}{N} \;, \\ \bar{P}_3 \;&=\; \frac{k_1k_2 + k_2l_3 + l_3l_1}{N} \;\;\text{with} \\ N &= k_1k_2 + k_2k_3 + k_3k_1 + k_1l_2 + k_2l_3 + k_3l_1 + l_1l_2 + l_2l_3 + l_3l_1. \end{aligned} \tag{4.152}$$

The third eigenvalue $\lambda_0 = 0$ indicates a linear dependence that is given by the conservation of probabilities: $P_1 + P_2 + P_3 = 1$. Apart from normalization, equation (4.152) is formally identical to equation (4.82).

Recording single enzyme trajectories provides information on steady states and on transient kinetics, which is commonly expressed in terms of transition probabilities: $P_{ij}(t, t+\Delta t) = P\big(\mathcal{X}(t+\Delta t) = \mathsf{S}_j \,|\, \mathcal{X}(t) = \mathsf{S}_i\big)$ denoted the probability that the enzymes is in state 'j' at time $t + \Delta t$ provided it was in state 'i' at time t. In the steady state of a Markov process transition probabilities are independent of time t: $P_{ij}(t, t+\Delta t) = P_{ij}(\Delta t)$ (section 3.1.3.3). The diagonal and off-diagonal stationary transition probabilities of the single molecule Michaelis-Menten mechanism are whereby we use indices modulo 3:

$$\begin{aligned}
P_{jj} =& -\frac{k_j + l_{j-1} + \lambda_2(1-\bar{P}_j)}{\lambda_1 - \lambda_2}\, e^{\lambda_1 \Delta t} + \\
&+ \frac{k_j + l_{j-1} + \lambda_1(1-\bar{P}_j)}{\lambda_1 - \lambda_2}\, e^{\lambda_2 \Delta t} + \bar{P}_j, \\
P_{j,j+1} =& \frac{k_j + \lambda_2\, \bar{P}_{j+1}}{\lambda_1 - \lambda_2}\, e^{\lambda_1 \Delta t} + \frac{k_j + \lambda_1\, \bar{P}_{j+1}}{\lambda_1 - \lambda_2}\, e^{\lambda_2 \Delta t} + \bar{P}_{j+1}, \\
P_{j,j-1} =& \frac{l_{j-1} + \lambda_2\, \bar{P}_{j-1}}{\lambda_1 - \lambda_2}\, e^{\lambda_1 \Delta t} + \frac{k_j + \lambda_1\, \bar{P}_{j-1}}{\lambda_1 - \lambda_2}\, e^{\lambda_2 \Delta t} + \bar{P}_{j-1}, \\
& j = 1, 2, 3, 1, 2, \ldots = (i \bmod 3) + 1\,,\ i = 0, 1, 2, 3, 4, \ldots\ .
\end{aligned} \tag{4.153}$$

It is straightforward to consider the transition probabilities for vanishing time intervals where we find as expected:

$$\lim_{\Delta t \to 0} P_{jj}(\Delta t) = 1 \quad \text{and} \quad \lim_{\Delta t \to 0} P_{j,j+1}(\Delta t) = \lim_{\Delta t \to 0} P_{j,j-1}(\Delta t) = 0\ ,$$

the off-diagonal elements corresponding to proper transitions converge to zero.

In the case of single enzyme kinetics the conversion of substrate into product can be modeled by a biassed continuous time random walk [378] on the finite lattice shown in figure 4.31.[58] Three states are characterized by the same number of substrate molecules: (n, C), (n, D), and (n, E), and hence the number of substrate molecules is a random variable $\mathcal{X}(t)$ with the probability

$$P_n(t) \;=\; P(\mathcal{X}(t) = n) \;=\; P_n^{(\mathbf{E})}(t) + P_n^{(\mathbf{C})}(t) + P_n^{(\mathbf{D})}(t)\ ,$$

where the superscripts refer to the state of the enzyme. In other words the three stochastic variables $\mathcal{X}_{\mathbf{E}}(t)$, $\mathcal{X}_{\mathbf{C}}(t)$, and $\mathcal{X}_{\mathbf{D}}(t)$ are lumped together in the variable $\mathcal{X}(t)$. As initial condition we assume $\mathcal{X}(0) = n_0$ substrate molecules and no product. A state of the system is fully characterized by n, the number of substrate molecules, and the state of the enzyme, **E**, **C** or **D**. Since we are dealing with a single enzyme molecule, the state space can be arranged as

[58] By *'biassed'* we express the fact that individual steps may have different weights.

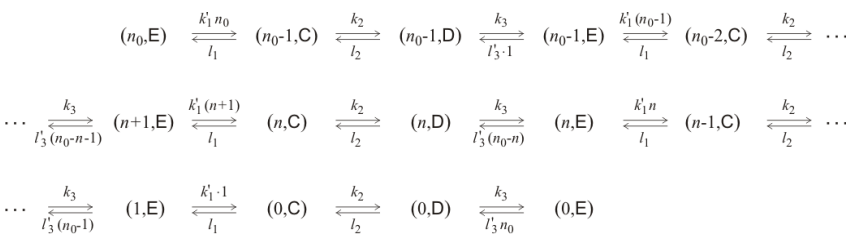


Fig. 4.31 Stochastic dynamics of substrate to product conversion [378]. The extended Michaelis-Menten mechanism (figure 4.2**A**) is applied to model the stochasticity of the substrate to product conversion. The single enzyme molecule occurs in three conformations: (i) free enzyme (**E**), (ii) enzyme bound to substrate (**C**), and (iii) enzyme bound to product (**D**). Because of the restriction to a single enzyme molecules the state space is a one dimensional array, and the stochastic model can be simplified to a (biassed) continuous time random walk.

a one dimensional lattice (see figure 4.31), and the stochastic process is a biassed continuous time random walk on this lattice. The bias is introduced by the transition probabilities, which differ for the individual transitions. Applying the same notation for the rate parameters as in the deterministic kinetic equation, we obtain for the probabilities:

$$\begin{aligned}
\frac{\mathrm{d}P_n^{(\mathbf{E})}(t)}{\mathrm{dt}} &= l_1 P_{n-1}^{(\mathbf{C})}(t) + k_3 P_n^{(\mathbf{D})}(t) - \big(k_1' n + l_3'(n_0 - n)\big) P_n^{(\mathbf{E})}(t), \\
\frac{\mathrm{d}P_n^{(\mathbf{C})}(t)}{\mathrm{dt}} &= k_1'(n+1) P_n^{(\mathbf{E})}(t) + l_2 P_n^{(\mathbf{D})}(t) - (k_2 + l_1) P_n^{(\mathbf{C})}(t)\,, \\
\frac{\mathrm{d}P_n^{(\mathbf{D})}(t)}{\mathrm{dt}} &= l_3'(n_0 - n) P_n^{(\mathbf{E})}(t) + k_2 P_n^{(\mathbf{C})}(t) - (k_3 + l_2) P_n^{(\mathbf{D})}(t)\,.
\end{aligned} \tag{4.154}$$

The equilibrium distribution of the probabilities is readily calculated and reported in the literature [378, 427]:

$$\begin{aligned}
\bar{P}_n^{(\mathbf{E})} &= \frac{n_0! K^{n_0-n}}{(n_0-n)!\, n!}\, \frac{1}{Q}\,, \\
\bar{P}_n^{(\mathbf{C})} &= \frac{n_0! K^{n_0-n-1}}{(n_0-n-1)!\, n!}\, K_1\, \frac{1}{Q}\,, \quad \text{and,} \\
\bar{P}_n^{(\mathbf{D})} &= \frac{n_0! K^{n_0-n-1}}{(n_0-n-1)!\, n!}\, K_1 K_2\, \frac{1}{Q}\,, \quad \text{with,} \\
K_1 &= \frac{k_1}{l_1}\,;\; K_2 = \frac{k_2}{l_2}\,;\; K_3 = \frac{k_3}{l_3}\,;\; K = K_1 K_2 K_3\,; \\
Q &= \frac{1 + K + n_0\,\kappa}{(1+K)^{n_0-1}} \quad \text{with} \quad \kappa = K_1(1+K_2)\,.
\end{aligned} \tag{4.155}$$

Expectation value and variance of the number of substrate molecules at equilibrium can be derived from the function Q:

$$\begin{aligned} \mathrm{E}(n) &= n_0 - \frac{\partial \ln Q(K)}{\partial \ln K} = \frac{n_0 + K}{1 + K} - \frac{K + n_0\,\kappa}{1 + K + n_0\,\kappa} \quad \text{and} \\ \mathrm{var}(n) &= \frac{\partial^2 \ln Q(K)}{\partial \ln K^2} = \frac{n_0\,K}{(1+K)^2} + \frac{K + n_0\,\kappa}{(1 + K + n_0\,\kappa)^2}\ . \end{aligned} \tag{4.156}$$

It is straightforward to calculate the values of the moments for large numbers of substrate molecules:

$$\mathrm{E}(n) = \frac{n_0}{1+K} \quad \text{and} \quad \mathrm{var}(n) = \frac{n_0\,K}{(1+K)^2} \quad \text{for large } n_0\ .$$

It is worth mentioning that precisely these expressions were obtained for the binomial distribution with the replacements $n \Leftrightarrow n_0$, $p \Leftrightarrow 1/(1+K)$, and $q \Leftrightarrow K/(1+K)$ in equation (2.36).

Extended Michaelis-Menten mechanism (scheme B*).* An alternative extension of the simple Michaelis-Menten mechanism that is suitable for handling single-molecule reactions (figure 4.2, scheme B) does not consider the enzyme-product complex $\mathbf{E}\cdot\mathbf{P}$ explicitly but instead introduces a conformational change of the enzyme molecule after product release, $\mathbf{E}^0 \to \mathbf{E}$, and this introduces a second linkage class:

$$\begin{gathered} \mathbf{S} + \mathbf{E} \underset{l_1}{\overset{k_1'}{\rightleftharpoons}} \mathbf{S}\cdot\mathbf{E} \underset{l_2'}{\overset{k_2}{\rightleftharpoons}} \mathbf{E}^0 + \mathbf{P} \quad \text{and} \\ \mathbf{E}^0 \underset{l_3}{\overset{k_3}{\rightleftharpoons}} \mathbf{E}\ . \end{gathered} \tag{4.157}$$

Again we use the *prime* in the notation in order to simplify writing of the rate parameters under pseudo-first order conditions: $k_1 = k_1' s_0$ and $l_2 = l_2' p_0$. Empirical evidence shows that the assumption of irreversible reaction steps 2 and 3 with $l_2 \approx 0$ and $l_3 \approx 0$ fits well the available data, and therefore the validity of the original Michaelis-Menten equation (4.14) is retained. In single molecule enzymology it is reasonable to assume that individual turnovers do not change substantially the substrate concentration, $[\mathbf{S}] = s_0$. Then, the linear ODE describing the probabilities of the single enzyme molecule to be in one of the three states has only two degrees of freedom because of the normalization relation $P_{\mathbf{E}} + P_{\mathbf{C}} + P_{\mathbf{E}^0} = 1$:

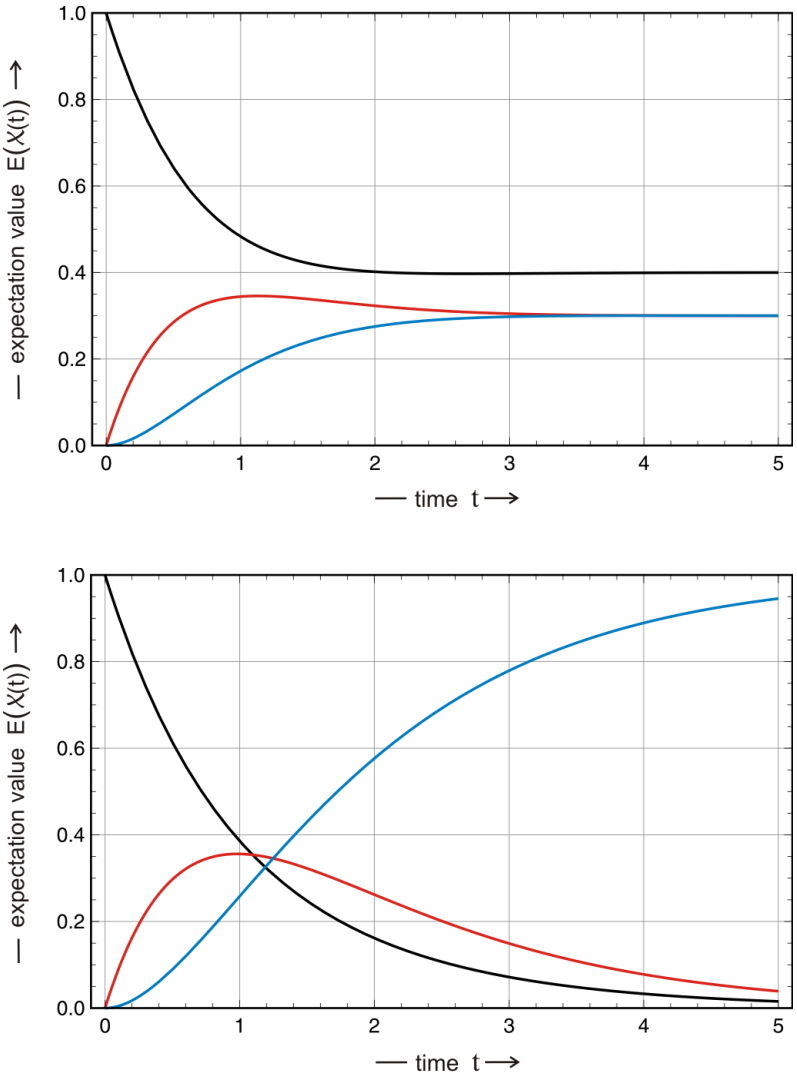


Fig. 4.32 Single enzyme turnover. The plots illustrate the single enzyme molecules mechanism (4.157). Since the stochastic variables are restricted to the values $\{0, 1\}$, the expectation value is identical to the probability $\mathrm{E}(\mathcal{X}) = \sum_n nP_n = P_1$ and equation (4.158) describes the evolution of the expectation values. In the upper plot we show the equilibration of the three variables in case of multiple turnovers. The lower plot concerns the completion of a single turnover that is achieved by setting $k_3 = 0$. The lower plot shows the integration of equation (4.158) with no recovery of the enzyme being tantamount to setting $k_3 = 0$. The enzyme cycle is arrested in state $\mathcal{X}_{\mathbf{E}_0} = 1$. Parameter choice: $k_1 = k_2 = 1$; upper plot: $k_3 = 1$, $l_1 = 0.3$; lower plot $k_3 = 0$, $l_1 = 0.1$. Color code: [**E**] black, [**C**] = [**S** · **E**] red, and [$\mathbf{E}_0$] blue.

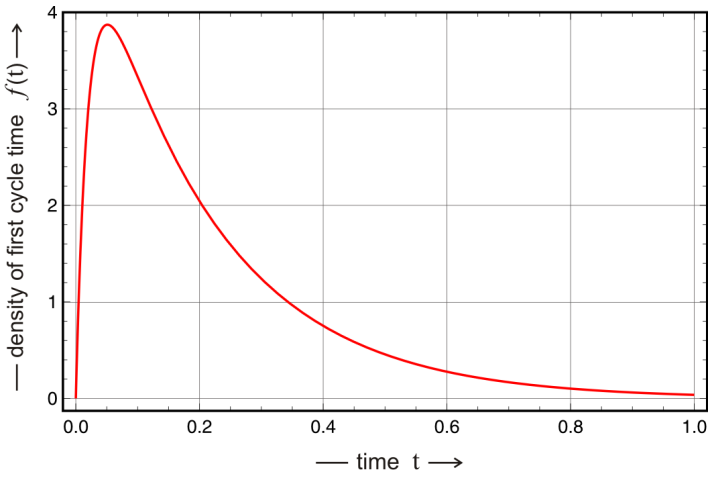


Fig. 4.33 Density of the first cycle waiting time $\mathcal{T}$. The plot shows the density of the time $\mathcal{T}$ required to complete the first turnover cycle, $\mathbf{E} \leftrightarrow \mathbf{C} \rightarrow \mathbf{E}^0 \rightarrow \mathbf{E}$, which is represented by the superposition of two exponential curves (4.160'):
$f(t) = \frac{\alpha\beta}{\beta-\alpha}\big(\exp(-\alpha t) - \exp(-\beta t)\big)$, with $\alpha = k_1[\mathbf{S}]$ and $\beta = k_2$.
This definition requires $\alpha \neq \beta$ and implies that the fast exponential is going up and the second one goes down since the denominator changes sign at $\alpha = \beta$.

$$\begin{aligned}
\frac{\mathrm{d}P_{\mathbf{E}}(t)}{\mathrm{dt}} &= +k_3\,P_{\mathbf{E}^0}(t) + l_1\,P_{\mathbf{C}}(t) - k_1\,P_{\mathbf{E}}(t)\ , \\
\frac{\mathrm{d}P_{\mathbf{C}}(t)}{\mathrm{dt}} &= +k_1\,P_{\mathbf{E}}(t) - (k_2 + l_1)\,P_{\mathbf{C}}(t)\ , \\
\frac{\mathrm{d}P_{\mathbf{E}^0}(t)}{\mathrm{dt}} &= +k_2\,P_{\mathbf{C}}(t) - k_3\,P_{\mathbf{E}^0}(t)\ .
\end{aligned} \tag{4.158}$$

Since the stochastic variables can adopt only two values, $\mathcal{X} \in \{0, 1\}$, the probabilities are identical with the expectation values:

$$\mathrm{E}\big(\mathcal{X}(t)\big) = 0 \cdot P_0(t) + 1 \cdot P_1(t) = P_1(t)\ .$$

Instead of dwelling further into the solutions of equation (4.158) we use it to study the first enzyme turnover cycle: With the assumption $k_3 = 0$ – no recovery of the enzyme – the equations are tailored for the calculation of a kind of first passage time, the time $\mathcal{T}$ that measures the time of the completion of the first turnover cycle. In other words $\mathcal{T}$ is the time when the enzyme molecule is the first time in conformation $\mathbf{E}^0$. Figure 4.32 shows solution curves of (4.158) without and with the assumption of vanishing k_3.

The first cycle completion time $\mathcal{T}$ is a random variable with the density $f_{\mathcal{T}}(t)$:

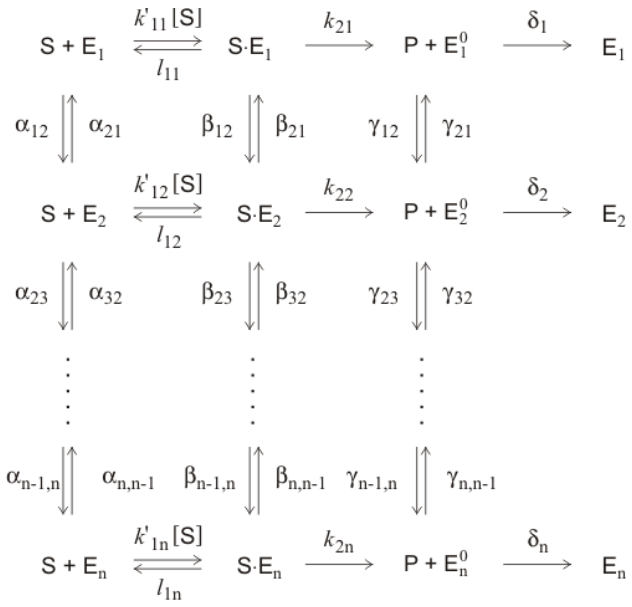


Fig. 4.34 A multistate model for enzyme reactions [262]. The extended Michaelis-Menten mechanism (figure 4.2, scheme B) is augmented by the assumption that the enzyme molecule can exist in a multitude of n distinct conformations, which differ in all their kinetic constants. The current theory of protein folding [354] predicts the existence of a multitude of hierarchically ordered conformations and single-molecule experiments are consistent with it.

$$\mathcal{T}: \quad f_{\mathcal{T}}(t)\ \mathrm{dt} \ = \ P(t < \mathcal{T} \leq t + \mathrm{dt} \ \text{ with } \ \int_0^\infty f_{\mathcal{T}}(t)\ \mathrm{dt} \ = \ 1\ . \qquad (4.159)$$

We can also interpret the density $f_{\mathcal{T}}(t)\Delta t$ as the probability that the enzyme molecule reaches the conformation $\mathbf{E}^0$ in the time interval between t and $t + \Delta t$, which can be easily calculated:

$$\Delta P_{\mathbf{E}^0}(\Delta t) \ = \ k_2\, P_{\mathbf{C}}\, \Delta t \ \text{ and } \ \lim \Delta t \to 0: \ \ f_{\mathcal{T}}(t) \ = \ \frac{\mathrm{d}P_{\mathbf{C}}}{\mathrm{dt}} \ = \ k_2\, P_{\mathbf{C}}(t)\,.$$

From the solution of equation (4.158) follows:

$$f_{\mathcal{T}}(t) \ = \ \frac{k_1[\mathbf{S}] \cdot k_2}{2d} \left(e^{(b+d)t} - e^{(b-d)t}\right)\ , \qquad (4.160)$$

the waiting time $f_{\mathcal{T}}(t)$ is a superposition of two exponential curves, a faster rising exponential and a slower decaying one (figure 4.33). In the limit of irreversibility of the first reaction, $\lim l_1 \to 0$, the mechanism is simple the sequence of two (pseudo) first order reaction steps and the waiting time

distribution becomes the convolution of the waiting times for the two individual steps: $f_{\mathcal{T}}(t) = (f_1 \otimes f_2)(t)$ or $f_{\mathcal{T}}(t) = \int_0^t \mathrm{d}\tau f_1(t-\tau) f_2(\tau)$. Provided $\mathbf{S} + \mathbf{E} \rightarrow \mathbf{S} \cdot \mathbf{E}$ and $\mathbf{S} \cdot \mathbf{E} \rightarrow \mathbf{E}^0 + \mathbf{P}$ are Poisson processes with $f_1 = k_1'[\mathbf{S}] \exp(-k_1'[\mathbf{S}]\, t)$ and $f_2 = k_2 \exp(-k_2 t)$ we obtain

$$f_{\mathcal{T}}(t) \;=\; \frac{k_1'[\mathbf{S}]\, k_2}{k_2 - k_1'[\mathbf{S}]} \left(e^{-k_1'[\mathbf{S}]\, t} - e^{-k_2 t} \right) \;, \qquad (4.160')$$

the faster exponential being the rising function and the slower exponential the decaying function.

The two extension of the simple Michaelis-Menten mechanism in a way remind of the debate on allosteric mechanisms of enzyme control in the nineteen sixtieth and seventieth. It had been developed for multimeric proteins[59] but the basic physical interpretation is essentially the same for a monomeric enzyme. The Koshland-Némethy-Filmer mechanism postulated by Daniel Koshland, and worked out for cooperative binding together with George Némethy and David Filmer of *induced fit* assumes that the protein changes shape on ligand binding. In other words, there is only a single conformation for the free protein molecules and all complexes with ligands have their specific protein structures. The Monod-Wyman-Changeaux mechanism [333] named after Jacques Monod, Jeffries Wyman, and Jean-Pierre Changeaux is based on the assumption that two or more enzyme conformations exist also in the absence of the ligand and ligand binding shifts the equilibrium between the states of the protein (For a more detailed but still short account on the Koshland and the Monod mechanism see [135, pp. 291-304]). Comparing with the two extensions of the extended Michaelis-Menten mechanism discussed here we recognize Koshland's principle of induced fit in scheme **A** and the Monod mechanism reflected by the two protein conformations $\mathbf{E}$ and $\mathbf{E}_0$ in scheme **B**. In contrast to binding in monomeric enzymes cooperative binding in multimeric proteins can lead to reaction rate profiles, $v([S])$, that are incompatible with the Michaelis-Menten mechanism [33, p. 267] or [34, p. 291] because they are `S´-shaped or sigmoid in some cases.

A multitude of distinct enzyme conformations differing in the rate parameters (figure 4.34) gives rise to *dynamical disorder*: The enzyme as well as the enzyme-substrate complex fluctuate randomly between n different conformational states but under a variety of conditions the form of the Michaelis-Menten equation is retained at the ensemble average level [262]. Some conditions of general relevance are:

(i) in the limit in which the interconversion rates of the enzyme substrate complexes, $\mathbf{S} \cdot \mathbf{E}_k$, are slower than the catalytic rates, $(\beta_{ij}) < (k_{2j})$ for $i, j = 1, \ldots, n$,

(ii) in the limit in which the interconversion rates between enzyme conformations are much larger than all other rates, and

[59] Multimeric proteins contain several, identical or different, subunits. The protein in the focus was hemoglobin that is a tetramer.

(iii) in the limit in which all Michaelis constants for individual reaction channels are practically the same, $(k_{21} + l_{11})/k_{11} \approx (k_{22} + l_{12})/k_{12} \approx \ldots \approx (k_{2n} + l_{1n})/k_{1n}$.

If condition (i) is fulfilled and the interconversion rates between the enzyme conformations are sufficiently small, the disorder becomes quasi-stationary and then the density of the waiting time can be approximated by a linear superposition of channel waiting times

$$f_{\mathcal{T}}(t) \;=\; \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i \frac{k_1[\mathbf{S}]\, k_2}{2d_i} \left(e^{(b_i+d_i)t} - e^{(b_i-d_i)t} \right) , \qquad (4.161)$$

where the coefficients w_i define the weights with which the individual channels contribute to the waiting time distribution in the ensemble.

Meanwhile dynamic disorder in enzyme reaction has been found and analyzed in an impressively large number of single-molecule experiments (see, e.g., [100, 101, 279, 442], and in addition the results obtained thereby fit well into the current theory of protein folding [151, 354]. Finally we mention that single-molecule enzymology shed new light on the mechanism of allosteric regulation of monomeric enzymes [199] and gave a clear hint that the different enzyme conformations exist also in absence of binding partners.

4.4.2 Fluorescence correlation spectroscopy

Correlation spectroscopy aims at measuring the fluctuations of a spectroscopic signal at thermodynamic equilibrium and consequently, its interpretation requires a theory of stochastic processes. Fluctuations are measured in a given volume V and by the $\sqrt{N}$-law the relative amplitude of fluctuations is the larger the smaller the number of molecules in the sample volume is, and this means small volumes and low concentrations facilitate the observation. In essence, fluorescence correlation spectroscopy (FCS) measures the numbers of molecules in a defined volume as a function of time. A few decades ago fluctuation spectroscopy was impossible since the signals were simply to weak. Two basic technical advances in fluorescence spectroscopy and microscopy made it possible to observe and evaluate fluorescence fluctuations: (i) application of high-power lasers and (ii) confocal microscopy. The spectacular improvement in laser technology raised the signal to noise ratio by several orders of magnitude and as said allows for the recording of signals of single molecules. The second advancement concerns the invention of the confocal microscope and provides the possibility of confining the molecule to be observed to very small volumes. Recordings of $\approx 1 \times 10^{-15}$ l became possible that correspond to cubes with an edge length of 1 μm. The autocorrelation function (section 3.1.4) is fairly easily accessible experimentally because technical devices called *autocorrelators* have been built [358], which

record directly the autocorrelation through data sampling of the process under investigation.

The quantity that is commonly derived from fluctuation measurements is a characteristic time, either a relaxation time of a chemical reaction, a relaxation time of a translational or a rotational diffusion process or the residence time of a flow in the volume of observation. The theoretical basis for the computation of rate parameters or diffusion coefficients from fluorescence correlation data is the fluctuation-dissipation theorem: The parameters, which determine the linear return to equilibrium of the system after a macroscopic perturbation are identical with the rates at which spontaneous fluctuations decay [265]. Originally, fluorescence correlation spectroscopy has been used to measure relaxation times of chemical reactions of the class $\mathbf{A} + \mathbf{B} \rightarrow \mathbf{C}$, in particular the binding of a fluorescent dye to a biomolecule, for example *ethidium bromide* to DNA [300]. Since the chemical reaction is almost always coupled to diffusion (see 'bimolecular reactive collisions') fluorescence correlation provides at the same time information on binding and diffusion.

The lower time limit for processes that can be observed by fluorescence is given by the rate of fluorescence excitation and emission of the photon. This basic photophysical process leads to the *antibunching term* in the autocorrelation function

$$G_F(\tau) \;=\; (1 \,-\, A_F \,\exp\,(-\tau/\tau_F) \;\; \text{with} \;\; A_F \,=\, 1 \;.$$

The excited state need not emit the fluorescence photon it can also undergo a transition to a non-fluorescent or *dark* triplet state and this yields another term in the autocorrelation function

$$G_T(\tau) \;=\; (1 \,+\, A_T \,\exp\,(-\tau/\tau_T) \;\; \text{with} \;\; A_T \,=\, \frac{\vartheta}{1-\vartheta} \;,$$

where ϑ is the fraction of molecules being *trapped* in the triplet state. Under commonly fulfilled conditions the relaxation times fulfill the conditions $\tau_{D,R} \gg \tau_T \gg \tau_F$, the autocorrelation function can be factorized in the sense of figure 4.8 and we obtain

$$G(\tau) \;=\; G_F(\tau) \times G_T(\tau) \times G_{R,D}(\tau) \;.$$

In other words on the characteristic timescales for fluorescence spectroscopy we need only consider the contributions of chemical reactions and diffusion where the diffusion time commonly sets the upper limit for the timescale of observable processes, and accordingly we shall write $G(\tau)$ for $G_{R,D}(\tau)$.

Theory of particle number correlations. The theory of concentration correlations has been developed in the nineteen sixtieth and seventieth for the application of inelastic light scattering to chemical kinetics of macromolecular kinetic reactions (see [44] and references therein). The adaptation of the theory to investigations of fluorescence correlation that, in essence, involves

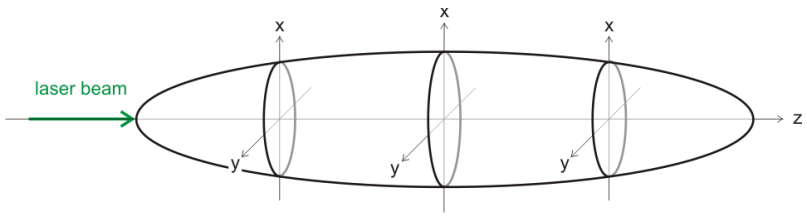


Fig. 4.35 Geometry of fluorescence measurements. A sketch of the beam waist in a fluorescence measurement with a confocal microscope. The active volume element containing the fluorescent sample in a prolate ellipsoid with a Gaussian intensity profile $I(\mathbf{r}) = I_0 \exp\bigl(-2(x^2+y^2)/w_{xy}^2 - 2z^2/w_z^2\bigr)$ with $\mathbf{r} = (x, y, z)$ and $w_z > w_{xy} = w_x = w_y$. The laser beam is oriented in the z-direction.

a drastic reduction of both the reaction volume and the concentration was done a few years later [12, 112]. In bulk solutions the four most important processes that can be analyzed by means of the autocorrelation function are: (i) directed laminar flow through the observation volume, (ii) translational diffusion, (iii) rotational diffusion, and (iv) chemical reaction. In addition, diffusion in natural and artificial membranes has been studied and *in vivo* measurements of molecules labelled with fluorescence markers were made in cells. Under the conditions of fluorescence correlation experiments diffusion is strongly coupled to chemical reactions and hence, fluctuation of concentrations in space and time have to be considered. The role of fluctuations is illustrated by means of a brief account of the theory [264].

The sample contains M different chemical species, $\mathbf{X}_1, \ldots, \mathbf{X}_M$, represented by random variables, $\mathcal{X}_1, \ldots, \mathcal{X}_M$ with the concentrations $x_1(\mathbf{r}, t), \ldots, x_M(\mathbf{r}, t)$, which are assumed to be functions of space and time. The sample is assumed to be at thermodynamic equilibrium: $\langle \mathcal{X}_j(\mathbf{r}, t) \rangle = x_j(\mathbf{r}, t) = \bar{x}_j(\mathbf{r}, t)$, and precisely $\bar{x}_j$ expresses the mean-square fluctuations of $\bar{x}_j(\mathbf{r}, t)$ in a unit volume following Poisson statistics. The fluctuations understood as deviations from the equilibrium values are denoted as in sections 4.1.1 and 4.1.3:

$$\chi_j(\mathbf{r}, t) \;=\; \delta x_j(\mathbf{r}, t) \;=\; x_j(\mathbf{r}, t) \;-\; \bar{x}_j \;,$$

and used as variables to describe the linear response to displacements from equilibrium. In systems combining diffusion and chemical reactions the fluctuations fulfill

$$\frac{\partial \chi_i(\mathbf{r}, t)}{\partial t} \;=\; D_i \, \nabla^2 \chi_i(\mathbf{r}, t) \;+\; \sum_{j=1}^{M} A_{ij} \chi_j(\mathbf{r}, t) \,;\; i = 1, \ldots, M \;, \qquad (4.162)$$

with D_i being the diffusion coefficient of $\mathbf{X}_i$ and $\mathrm{A} = \{A_{ij}\}$ being the relaxation matrix (4.36).

Next the correlations between concentrations have to expressed in terms of the solutions $\chi_i(\mathbf{r},t)$ of this linear PDE for the appropriate boundary conditions given by the geometry of the fluorescence experiment (figure 4.35) and initial conditions $\chi_j(\mathbf{r},0)$ with $j=1,\ldots,M$. The correlation function of the concentrations of two chemical species $\mathbf{X}_j$ and $\mathbf{X}_l$ at times t and $t+\tau$ and positions $\mathbf{r}_1$ and $\mathbf{r}_2$,

$$C_{j\,l}(\mathbf{r}_1,\mathbf{r}_2,\tau) \;=\; \langle \chi_j(\mathbf{r}_1,t)\,\chi_l(\mathbf{r}_2,t+\tau)\rangle \;,$$

measures the probability to find a molecule of species $\mathbf{X}_j$ at position $\mathbf{r}_1$ at time t and a molecule of species $\mathbf{X}_l$ at position $\mathbf{r}_2$ a time interval $\Delta t=\tau$ later. Three conditions are assumed to be fulfilled: (i) microscopic reversibility,

$$C_{j\,l}(\mathbf{r}_1,\mathbf{r}_2,\tau) \;=\; C_{j\,l}(\mathbf{r}_1,\mathbf{r}_2,-\tau) \;=\; C_{j\,l}(\mathbf{r}_1,\mathbf{r}_2,|\tau|) \;,$$

(ii) strong or at least weak stationarity,

$$C_{j\,l}(\mathbf{r}_1,\mathbf{r}_2,\tau) \;=\; \langle \chi_j(\mathbf{r}_1,0)\,\chi_l(\mathbf{r}_2,\tau)\rangle \;, \quad \text{and} \tag{4.163}$$

(iii) lack of zero-time correlations between the positions of different molecules no matter whether they belong to the same or to different species:

$$\langle \chi_j(\mathbf{r}_1,0)\,\chi_k(\mathbf{r}_3,0)\rangle \;=\; \bar{x}_j\,\delta_{jk}\,\delta(\mathbf{r}_1-\mathbf{r}_3) \;. \tag{4.164}$$

Condition (iii) is fulfilled in ideal chemical solutions where there is no interaction between molecules except the collisions discussed in section 4.1.4. In other words the correlation lengths are much smaller than the distances between particles.

Solutions $\chi_j(\mathbf{r},t)$ with the initial conditions $\chi_j(\mathbf{r},0)$ are derived by means of spatial Fourier transform. Inserting the expressions for the transformed derivatives

$$\mathcal{F}\left(\frac{\partial u}{\partial x}\right) \;=\; -i\!\!i\,q_x\,\mathcal{F}(u) \quad \text{and} \quad \mathcal{F}\left(\frac{\partial^2 u}{\partial x^2}\right) \;=\; -\,q_x^2\,\mathcal{F}(u)$$

into equation (4.162) yields a linear ODE that can be readily solved by means of an eigenvalue problem (see section 4.1.3):

$$\frac{\mathrm{d}\hat{x}_i(\mathbf{q},t)}{\mathrm{dt}} \;=\; \sum_{j=1}^{M} R_{ij}\,\hat{x}_j(\mathbf{q},t) \quad \text{with} \quad \mathrm{R} \;=\; \{R_{ij}=A_{ij}-D_i\,q^2\,\delta_{ij}\} \;, \tag{4.165}$$

with the Fourier transform of the concentrations defined by

$$\hat{x}_i(\mathbf{q},t) \;=\; \frac{1}{(2\pi)^{3/2}}\int_{-\infty}^{\infty} \mathrm{d}\mathbf{r}\,e^{i\!\!i\,\mathbf{q}\cdot\mathbf{r}}\,\chi_i(\mathbf{r},t) \;.$$

Diagonalization of the reaction-diffusion matrix, $\Lambda = \mathrm{B}^{-1}\cdot\mathrm{R}\cdot\mathrm{B}$ with $\mathrm{B} = \{b_{ij}\}$ and $\mathrm{B}^{-1} = \mathrm{H} = \{h_{ij}\}$ yields the solution in frequency space

$$\hat{x}_i(\mathbf{q},t) \;=\; \sum_{k=1}^{M} b_{ik}\,\beta_k(0)\,e^{\lambda_k\,t} \quad \text{with} \quad \beta_k(0) = \sum_{j=1}^{M} h_{kj}\,\hat{x}_j(\mathbf{q},0)\;.$$

Insertion in equation (4.163) and exchange of Fourier transform and ensemble average yields

$$\begin{aligned}
\langle\chi_j(\mathbf{r}_1,0)\chi_l(\mathbf{r}_2,\tau)\rangle \;&=\; \frac{1}{(2\pi)^{3/2}}\int_{-\infty}^{\infty}\mathrm{d}\mathbf{q}\,e^{-i\mathbf{q}\cdot\mathbf{r}}\,\langle\chi_j(\mathbf{r}_1,0)\,\hat{x}_l(\mathbf{q},\tau)\rangle \;=\\
&=\; \frac{1}{(2\pi)^{3/2}}\int_{-\infty}^{\infty}\mathrm{d}\mathbf{q}\,e^{-i\mathbf{q}\cdot\mathbf{r}}\sum_{k=1}^{M} b_{lk}\,e^{\lambda_k\,\tau)}\sum_{i=1}^{M} h_{ki}\;\langle\chi_j(\mathbf{r}_1,0)\,\hat{x}_i(\mathbf{q},0)\rangle \;=\\
&=\; \frac{1}{(2\pi)^{3}}\int_{-\infty}^{\infty}\mathrm{d}\mathbf{q}\,e^{-i\mathbf{q}\cdot\mathbf{r}}\sum_{k=1}^{M} b_{lk}\,e^{\lambda_k\,\tau)}\sum_{i=1}^{M} h_{ki}\int_{-\infty}^{\infty}\mathrm{d}\mathbf{r}\,e^{i\mathbf{q}\cdot\mathbf{r}}\;\langle\chi_j(\mathbf{r}_1,0)\,\chi_i(\mathbf{r}_3,0)\rangle\;.
\end{aligned}$$

Making use now of equation (4.164) we get the final result:

$$\begin{aligned}
C_{jl}(\mathbf{r}_1,\mathbf{r}_2,\tau) \;&=\; \langle\chi_j(\mathbf{r}_1,0)\,\chi_l(\mathbf{r}_2,\tau)\rangle \;=\\
&=\; \frac{1}{(2\pi)^3}\,\bar{x}_j\int_{-\infty}^{\infty}\mathrm{d}\mathbf{q}\,e^{i\mathbf{q}\cdot(\mathbf{r}_1-\mathbf{r}_2)}\sum_{k=1}^{M} b_{lk}\,h_{kj}\;\exp\left(\lambda_k\,\tau\right))\;.
\end{aligned} \tag{4.166}$$

It is easily verified that the correlation function fulfils the expected symmetry property $C_{j\,l}(\mathbf{r}_1,\mathbf{r}_2,\tau) = C_{lj}(\mathbf{r}_1,\mathbf{r}_2,\tau)$ and $C_{j\,l}(\mathbf{r}_1,\mathbf{r}_2,\tau) = C_{j\,l}(\mathbf{r}_2,\mathbf{r}_1,\tau)$. The correlation function is proportional to the equilibrium concentration and decreases with increasing time delay, $\Delta t = \tau$, since the eigenvalues $\lambda_k = -\tau_{R_k}^{-1}$ of the relaxation matrix are negative: In particular, the eigenvalues for diffusion are always negative, $\lambda = -D\,q^2$, the same is essentially true for chemical reactions, where some but never all eigenvalues might be zero. For vanishing delay the autocorrelation function becomes a Dirac delta-function as expected: $\lim_{\tau\to 0} C_{jj}(\mathbf{r}_1,\mathbf{r}_2,0) = \bar{x}_j\delta(\mathbf{r}_1-\mathbf{r}_2)$ (4.164).

Fluorescence correlation measurements. The quantity measured in fluorescence experiments is the number of photons $n(t)$ emitted by the sample and collected in the detector

$$n(t) \;=\; \Delta t\int\mathrm{d}\mathbf{r}\,I(\mathbf{r})\sum_{i=1}^{M} Q_i x_i(\mathbf{r},t)\;.$$

Herein, $I(\mathbf{r})$ is the distribution of the light performing the excitation of the sample, and Q_i is the specific molecular parameter consisting of two factors, (i) the absorption cross section and (ii) the fluorescence quantum yield of molecules $\mathbf{X}_i$. Then the fluctuation in the photon count is

$$\delta n(t) \;=\; n(t) \,-\, \bar{n} \;=\; \delta t \int_{-\infty}^{\infty} \mathrm{d}\mathbf{r}\, I(\mathbf{r}) \sum_{i=1}^{M} Q_i\, \chi_i(\mathbf{r},t)\;, \tag{4.167}$$

and its average or equilibrium value is obtained by Fourier transform and integration:

$$\bar{n} \;=\; \Delta t \int \mathrm{d}\mathbf{r}\, I(\mathbf{r}) \sum_{i=1}^{M} Q_i \bar{x}_i(\mathbf{r},t) \;=\; (2\pi)^{3/2}\, \hat{I}(0)\, \Delta t \sum_{i=1}^{M} Q_i x_i(\mathbf{r},t)\;,$$

where $\hat{I}(\mathbf{q}) = (2\pi)^{-3/2} \int \mathrm{d}\mathbf{r}\, e^{-i\mathbf{q}\cdot\mathbf{r}} I(\mathbf{r})$. Making use of the ergodicity of the system we can write the fluorescence autocorrelation function as

$$\begin{aligned} G(\tau) \;&=\; \frac{1}{\bar{n}^2}\, \langle \delta n(0)\, \delta n(\tau) \rangle \;= \\ &=\; \frac{(\Delta t)^2}{\bar{n}^2} \int \mathrm{d}\mathbf{r}_1\, I(\mathbf{r}_1) \int \mathrm{d}\mathbf{r}_2\, I(\mathbf{r}_2) \sum_{j,l} Q_j Q_l\, \langle \chi_j(\mathbf{r}_1,0) \chi_l(\mathbf{r}_2) \rangle \;= \\ &=\; \frac{(\Delta t)^2}{\bar{n}^2} \int \mathrm{d}\mathbf{q}\, |I(\mathbf{q})|^2 \sum_{j,l} Q_j Q_l \bar{x}_j \sum_{k=1}^{M} b_{lk} h_{kj}\, e^{\lambda_k \tau}\;. \end{aligned}$$

The expression is completed by inserting a Gaussian intensity profile for the illumination of the sample (figure 4.35):

$$I(\mathbf{r}) \;=\; I_0\, \exp\left(-\,\frac{2\,(x^2+y^2)}{w_{xy}^2} \;-\; \frac{2z^2}{w_z^2} \right)\;, \tag{4.168}$$

which has the shape of a prolate ellipsoid with the shorter axes in the x- and y-direction and the longer axis in the z-direction, $w_x = w_y = w_{xy} < w_z$, and $\omega = w_z / w_{xy} \gg 1$. Fourier transformation yields

$$I(\mathbf{q}) \;=\; \frac{I_0\, w_{xy}\, w_z}{8} \exp\left(-\,\frac{w_{xy}^2}{8}\,(q_x^2+q_y^2) \;-\; \frac{w_z^2}{8} z^2 \right)\;,$$

and eventually we obtain the final equation for the autocorrelation function:

$$\begin{aligned} G(\tau) \;&=\; \frac{1}{(2\pi)^3} \frac{1}{\left(\sum_{i=1}^{M} Q_i \bar{x}_i\right)^2} \cdot \\ &\cdot \int_{-\infty}^{\infty} \mathrm{d}\mathbf{q}\, \exp\left(-\frac{w_{xy}^2}{4}(q_x^2+q_y^2) \;-\; \frac{w_z^2}{4} q_z^2 \right) \sum_{j=1}^{M} \sum_{l=1}^{M} Q_j Q_l \bar{x}_j \sum_{k=1}^{M} b_{lk} h_{kj}\, e^{\lambda_k \tau}\;. \end{aligned} \tag{4.169}$$

We remark that according to (4.165) the eigenvalues λ_k and the eigenvectors depend on $\mathbf{q}$ and for each particular case the $\mathbf{q}$-dependence has to be calculated from the relaxation dynamics.

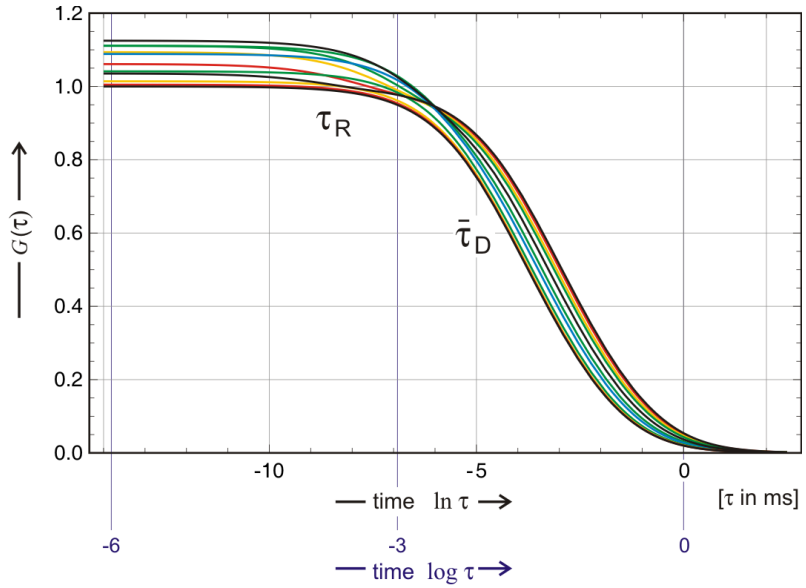


Fig. 4.36 Inclusion complexes of pyronines in cyclodextrin [6]. The autocorrelation curves $G(\tau)$ were calculated from equation (4.173) with the parameters given in [6]: $N_{\mathbf{G}} + N_{\mathbf{C}} = 1$, $\tau_{\mathbf{G}} = 0.25\,\mathrm{ms}$, $\tau_{\mathbf{C}} = 0.60\,\mathrm{ms}$, $\omega = 5$, $K = 2\,\mathrm{mM}^{-1}$, $Q = 0.5$, and $h = 500\,\mathrm{ms}^{-1}$, and the cyclodextrin concentrations $[\mathbf{H}]_0$ were: 12 (black), 6 (red), 3 (yellow), 2 (green), 1 (black), 0.5 (green), 0.3 (blue), 0.1 (green), 0.03 (yellow), 0.01 (red), and 0 mM (black).

Examples of fluorescence correlations. The simplest conceivable example analyzed by fluorescence correlation is the diffusion of a single chemical species **X** with the concentration $[\mathbf{X}] = x(\mathbf{r},t)$ and $\chi(\mathbf{r},t) = x(\mathbf{r},t) - \bar{x}$. Equation (4.162) becomes a simple diffusion equation

$$\frac{\partial \chi(\mathbf{r},t)}{\partial t} \;=\; D\,\nabla^2 \chi(\mathbf{r},t) \quad \text{and} \quad \hat{\chi}(\mathbf{q},t) \;=\; \hat{\chi}(\mathbf{q},0)\exp\left(-Dq^2\,t\right) .$$

The single eigenvalue of matrix R is $\lambda = -Dq^2$, the eigenvector is trivially $b = h = 1$ and insertion into (4.169) yields

$$\begin{aligned} G(\tau) \;&=\; \frac{(2\pi)^{-3}}{Q^2\bar{x}^2} \int_{-\infty}^{\infty} \mathrm{d}\mathbf{q}\, Q^2 \bar{x}\, e^{-\left(\frac{w_{xy}^2}{4}(q_x^2+q_y^2)+\frac{w_z^2}{4}q_z^2+D(q_x^2+q_y^2+q_z^2)t\right)} \;= \\ &=\; \frac{1}{\bar{N}}\left(1+\frac{\tau}{\tau_D}\right)^{-1}\left(1+\frac{\tau}{\omega^2\tau_D}\right)^{-1/2} \quad \text{with} \quad \bar{N} = \bar{x}\,V\,. \end{aligned} \tag{4.170}$$

Herein $\bar{N}$ is the number of molecules $\mathbf{X}$ in the effective sampling volume $V = \pi^{3/2} w_{xy}^2 w_z$, $\tau_D = w_{xy}^2/(4D)$ is the characteristic diffusion time across the illuminated ellipsoid, and $\omega^2 \tau_D = w_z^2/(4D)$ is the diffusion time along the ellipsoid. Each degree of freedom in diffusion contributes a factor $(1 - \tau/\tau'_D)^{-1/2}$ where $\tau'_D = (\omega')^2 \tau_D$ and ω' is a factor depending on the geometry of the illuminated volume. For an extended prolate ellipsoid we have $w_z \gg w_{xy}$ and then the autocorrelation function for diffusion in two dimensions

$$G(\tau) \;=\; \frac{1}{\bar{N}} \left(1 + \frac{\tau}{\tau_D}\right)^{-1} \tag{4.171}$$

is also a good approximation for the three-dimensional case: The relaxation of the fluctuation of the number of molecules in the sampling volume is approximately determined by the diffusion in the smaller dimensions. Recording of the autocorrelation function provides two results: (i) $G(0) = \bar{N}_{\mathbf{X}}^{-1}$, the number of particles in the beam waist, and (ii) $D = w_{xy}^2/4\tau_D$ the translational diffusion coefficient of $\mathbf{X}$.

The extension to M diffusing chemical species, $\mathbf{X}_1, \ldots, \mathbf{X}_M$, is straightforward [264]:

$$G(\tau) \;=\; \frac{1}{\left(\sum_{i=1}^{M} Q_i \bar{N}_i\right)^2} \sum_{j=1}^{M} Q_j^2 \bar{N}_j \left(1 + \frac{\tau}{\tau_{D_j}}\right)^{-1} \left(1 + \frac{\tau}{\omega^2\, \tau_{D_j}}\right)^{-1/2} , \tag{4.172}$$

the amplitude of the contribution of each species is weighted by its fluorescence quantum yield Q_j, $\bar{N}_j$ is the mean number of molecules $\mathbf{X}_j$ in the beam waist, and D_j is its diffusion coefficient.

The coupling between translational diffusion and chemical reactions leads to a more complex expression for the autocorrelation function. An excellent theoretical and experimental treatment is found in the literature [6]: the formation of inclusion complexes of pyronines with cyclodextrin,

$$\mathbf{G} + \mathbf{H} \underset{h}{\overset{k}{\rightleftharpoons}} \mathbf{C} \;\text{ with }\; K = k/h \;.$$

The fluorescent guest molecule $\mathbf{G}$ binds to the non-fluorescent host $\mathbf{H}$ and forms a fluorescent inclusion complex $\mathbf{C}$. Conditions are chosen under which the host concentration is much larger than the guest concentration: $[\mathbf{H}_0] \approx [\mathbf{H}] \gg [\mathbf{G}]$. It is useful to introduce a mean diffusion time $\bar{\tau}_D$, which is calculated from a weighted mean diffusion coefficient:

$$\bar{\tau}_D \;=\; \frac{w_{xy}^2}{4\bar{D}} \;\text{ with }\; \bar{D} \;=\; x_{\mathbf{G}} D_{\mathbf{G}} + x_{\mathbf{H}} D_{\mathbf{H}} \;,$$

where $x_{\mathbf{G}} = N_{\mathbf{G}}/(N_{\mathbf{G}} + N_{\mathbf{C}})$ and $x_{\mathbf{C}} = N_{\mathbf{C}}/(N_{\mathbf{G}} + N_{\mathbf{C}})$. Then the autocorrelation function is of the form

$$G_R(\tau) = \frac{1}{N_{\mathbf{G}} + N_{\mathbf{C}}} \left(1 + \frac{\tau}{\bar{\tau}_D}\right)^{-1} \left(1 + \frac{\tau}{\omega^2 \bar{\tau}_D}\right)^{-1/2} \left(1 + A_R e^{-\frac{\tau}{\tau_R}}\right), \quad (4.173)$$

where relaxation amplitude and relaxation time are given by:

$$A_R([\mathbf{H}]) = \frac{N_{\mathbf{G}} N_{\mathbf{C}} (Q_{\mathbf{G}} - Q_{\mathbf{C}})^2}{(Q_{\mathbf{G}} N_{\mathbf{G}} + Q_{\mathbf{C}} N_{\mathbf{C}})^2} = \frac{K[\mathbf{H}]_0 (1-Q)^2}{(1 + QK[\mathbf{H}]_0)^2} \quad \text{and}$$

$$\tau_R([\mathbf{H}]) = \left(h(1 + K[\mathbf{H}]_0)\right)^{-1} \quad \text{with} \quad Q = \frac{Q_{\mathbf{C}}}{Q_{\mathbf{G}}}, \; K[\mathbf{H}]_0 = \frac{N_{\mathbf{C}}}{N_{\mathbf{G}}}.$$

The relaxation curves were recalculated with the parameter values given in [6] and the result is shown in figure 4.36. The family of curves calculated for different value of total cyclodextrin concentration, $[\mathbf{H}]_0$, shows two relaxation processes, the faster one corresponding to the association reaction with a relaxation time τ_R and the slower process caused by diffusion of the two fluorescent species, the guest molecule **G** and the inclusion complex **C**. The amplitude of the chemical relaxation process $A_R([\mathbf{H}])$ increases first with increasing cyclodextrin concentration, $A_R([\mathbf{H}]) \approx K[\mathbf{H}]_0(1-Q)^2$ for small values of $[\mathbf{H}]$, passes a maximum at $[\mathbf{H}]_0 = 1/(QK)$, and then decreases according to $A_R([\mathbf{H}]) \approx (1-Q)^2/(Q^2[\mathbf{H}]_0)$ for large $[\mathbf{H}]_0$ values. Coupling of the chemical reaction with the diffusion process gives results in non-monotonous dependence of the relaxation amplitude on the host concentrations.

Provided parameter estimation can be done successfully (see section 4.1.5) fluorescence correlation spectroscopy allows for the determination of otherwise hard to obtain data: (i) the local concentration in the beam waist through $G(0)$, (ii) local translational diffusion coefficients from diffusion relaxation times, $\tau_D = w_{xy}^2/(4D)$, and (iii) relaxation times of chemical reactions, τ_R. Rotational diffusion constants can also be derived from fluorescence correlation [103, 459] and provide direct information on the size of molecules. In particular, the formation of molecular aggregates can be detected by determining the molecular radius. Technical advances in laser technique and microscopes allowed for a dramatic increase in resolution such that autocorrelation data from single molecules can be detected now [285, 319, 381].

4.5 Scaling and size expansions

Master equations when applied to real world chemical systems encounter serious limitations with respect to solvability, both analytic and numerical. As we have seen the analytical approach becomes extremely sophisticated already for single step bimolecular reactions (section 4.3.3) and numerical simulations cannot be carried out with reasonable resources when particle numbers become large (see section 4.6). In contrast Fokker-Planck and stochastic differential equations are much easier to handle and accessible to upscaling. In the section dealing with chemical Langevin equations (section 4.2.3) we discussed the approximating assumptions that allow for a transition from discrete particle numbers to continuous concentrations.

In this section we shall discuss ways to relate chemical master equations to Fokker-Planck equations. In particular, we shall solve master equations through approximation methods based on expansions in suitable parameters as we mentioned one case already in section 4.2.1 in form of the expansion of master equations in Taylor series with the jump moments being the coefficients. Truncation after the second term yields a Fokker-Planck equation. It is important to note that every diffusion process can be approximated by a jump process but the reverse is not true: Similarly as we saw in case of the transition from master to Langevin equations, there are master equations for which no approximation by a Fokker-Planck equation exists. A particularly useful expansion technique based on system sizes has been introduced by Nicholas van Kampen [438, 439]. This expansion method can be used, for example, to handle and discuss fluctuations without calculating solutions with full population sizes.

4.5.1 Kramers-Moyal expansion

The two physicists Hendrik Anthony Kramers and José Enrique Moyal proposed a general expansion of master equations, which is a kind of Taylor expansion in jump moments (section 3.2.3.1) applied to the integral equivalent of the master equation[60]

$$\frac{\partial P(x,t)}{\partial t} \;=\; \int \mathrm{d}z \left(W(x|z,t)\, P(z,t) \;-\; W(z|x,t)\, P(x,t) \right), \tag{4.174}$$

Starting point is the probability of the transition from the probability density at time t to the probability density at time $t+\tau$:

$$P(x,t+\tau) \;=\; \int \mathrm{d}z\, W(x,t+\tau|z,t)\, P(z,t)\ . \tag{4.175}$$

[60] A comprehensive presentation of different ways to derive series expansions leading to Fokker-Planck equation is found in [383, pp. 63-76].

We aim at the derivation of an expression for the differential $\mathrm{d}P$, which requires knowledge of the transition probabilities $W(x, t+\tau|z, t)$ – at least for small τ – and knowledge of the jump moments $\alpha_n(z, t, \tau)$:

$$\begin{aligned} \alpha_n(z,t,\tau) &= \left\langle \Big(\mathcal{X}(t+\tau) - \mathcal{X}(t)\Big)^n \right\rangle \Big|_{\mathcal{X}(t)=z} = \\ &= \int \mathrm{d}x\, (x-z)^n\, W(x, t+\tau|z,t)\ . \end{aligned} \tag{4.176}$$

Here $\mathcal{X}(t) = z$ implies that the random variable $\mathcal{X}(t)$ adopts the sharp value z at time t. Next we introduce $z = x - \Delta x$ into the integrand in equation (4.174) and expand a Taylor series in Δx around the value $x + \Delta x$:

$$\begin{aligned} &W(x, t+\tau|z,t)\, P(z,t) \ = \\ &\quad = \ W\big((x-\Delta x) + \Delta x, t+\tau|(x-\Delta x), t\big)\, P\big((x-\Delta x), t\big) \ = \\ &\quad = \ \sum_{n=0}^{\infty} \frac{(-\Delta x)^n}{n!}\, \frac{\partial^n}{\partial x^n}\Big(W(x+\Delta x, t+\tau|x,t)\, P(x,t)\Big)\ . \end{aligned}$$

Insertion into (4.175) and integration yields

$$\begin{aligned} P(x,t+\tau) \ &= \ \int \mathrm{d}(\Delta x) \sum_{n=0}^{\infty} \frac{(-\Delta x)^n}{n!}\, \frac{\partial^n}{\partial x^n}\Big(W(x+\Delta x, t+\tau|x,t)\, P(x,t)\Big) \ = \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!}\, \frac{\partial^n}{\partial x^n} \int \mathrm{d}(\Delta x)\, (\Delta x)^n \Big(W(x+\Delta x, t+\tau|x,t)\, P(x,t)\Big) \ = \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!}\, \frac{\partial^n}{\partial x^n}\, \alpha_n(x,t,\tau)\, P(x,t)\ . \end{aligned}$$

For the derivation of a convenient expression we perform a Taylor expansion of the jump moments

$$\frac{\alpha_n(x,t,\tau)}{n!} \ = \ \sum_{k=0}^{\infty} \frac{\tau^k}{k!}\, \Theta_k^{(n)} \quad \text{with} \quad \Theta_k^{(n)} \ = \ \frac{1}{n!}\, \frac{\partial^k \alpha_n}{\partial \tau^k}\ .$$

and truncate after the linear term in τ. Since $\Theta_0^{(n)}$ has to vanish because the transition probability fulfils the initial condition $W(x, t|x-\Delta x, t) = \delta(\Delta x)$ we find,

$$\frac{\alpha_n(x,t,\tau)}{n!} \ = \ \Theta_1^{(n)}\, \tau \ + \ \mathcal{O}(\tau^2)\ ,$$

with the linear term baring the only nonzero coefficient. Therefore we can drop the subscript, $\Theta^{(n)} \equiv \Theta_1^{(n)}$, move the term with $n = 0$ to the l.h.s., and divide by τ:

$$\frac{P(x,t+\tau)-P(x,t)}{\tau} \;=\; \sum_{n=1}^{\infty}(-1)^n\,\frac{\partial^n}{\partial x^n}\Big(\Theta^{(n)}\,P(x,t)\Big)\,.$$

Performing the limit $\tau \to 0$ finally yields the expansion of the master equation

$$\frac{\partial P(x,t)}{\partial t} \;=\sum_{n=1}^{\infty}(-1)^n\,\frac{\partial^n}{\partial x^n}\Big(\Theta^{(n)}\,P(x,t)\Big)\;.$$

We remark that the above given derivation corresponds to a forward stochastic process and in addition to this forward there exists also a backward Kramers-Moyal expansion.

Assuming explicit time independence of the transition matrix and the jump moments we obtain the conventional form of the Kramers-Moyal expansion

$$\begin{aligned}\frac{\partial P(x,t)}{\partial t} \;&=\sum_{n=1}^{\infty}\frac{(-1)^n}{n!}\,\frac{\partial^n}{\partial x^n}\Big(\alpha_n(x)\,P(x,t)\Big) \quad \text{with}\\ \alpha_n(x) \;&=\; \int_{-\infty}^{\infty}(z-x)^n\;W(x,z-x)\,\mathrm{d}z\;.\end{aligned} \tag{4.177}$$

In case the Kramers-Moyal expansion is terminated after the second term the result is a Fokker-Planck equation of the form:

$$\frac{\partial P(x,t)}{\partial t} \;=\; -\,\frac{\partial}{\partial x}\Big(\alpha_{\,1}(x)\,P(x,t)\Big)\;+\;\frac{1}{2}\,\frac{\partial^2}{\partial x^2}\Big(\alpha_{\,2}(x)\,P(x,t)\Big)\;. \tag{4.178}$$

The two jump moments represent the conventional drift and diffusion terms: $\alpha_1(x) \equiv A(x)$ and $\alpha_2(x) \equiv B(x)$. We remark that we used nowhere the condition of a one-step birth and death processes and therefore (4.178) is generally valid.

4.5.2 Small noise expansion

For large particle numbers noise fulfilling a $\sqrt{N}$-law may be very small and advantage of this fact can be made in *small noise expansions* of stochastic differential and Fokker-Planck equations. Then the SDE can be written as:

$$\mathrm{d}x \;=\; a(x)\;\mathrm{dt}\;+\;\varepsilon\,b(x)\;\mathrm{d}W(t)\;, \tag{4.179a}$$

where the solution is assumed to be of the form

$$x_\varepsilon(t) \;=\; x_0(t)\;+\;\varepsilon\,x_1(t)\;+\;\varepsilon^2\,x_2(t)\;+\;\ldots \tag{4.179b}$$

Solutions can be derived term by term and $x_0(t)$, for example, is the solution of the deterministic differential equation, $\mathrm{d}x = a(x)\ \mathrm{dt}$ with the initial condition $x_0(0) = c_0$.

In the small noise limit a suitable Fokker-Planck equation is of the form

$$\frac{\partial P(x,t)}{\partial t} \;=\; -\frac{\partial}{\partial x}\Big(A(x)\,P(x,t)\Big) \;+\; \frac{1}{2}\,\varepsilon^2\,\frac{\partial^2}{\partial x^2}\Big(B(x)\,P(x,t)\Big)\;, \qquad (4.180a)$$

where the variable x and the probability $P(x,t)$ density are scaled

$$\xi \;=\; \frac{x - x_0(t)}{\varepsilon} \quad \text{and} \quad P_\varepsilon(\xi,t) \;=\; \varepsilon\,P(x,t|c_0,0)\;, \qquad (4.180b)$$

and the probability density is assumed to be of the form

$$P_\varepsilon(\xi,t) \;=\; P_\varepsilon^{(0)}(\xi,t) \;+\; \varepsilon\,P_\varepsilon^{(1)}(\xi,t) \;+\; \varepsilon^2\,P_\varepsilon^{(2)}(\xi,t) \;+\; \ldots \qquad (4.180c)$$

This innocent looking approach has to face two problems: (i) There is no guarantee that the two expansion series (4.179b) and (4.180c) converge, and (ii) explicit calculations based on the series expansions are commonly quite sophisticated [157, pp.169-184].

For the purpose of illustration we consider one special example, the Ornstein-Uhlenbeck process, which is exactly solvable (see section 3.2.2.3). The stochastic differential equations is of the form

$$\mathrm{d}x \;=\; -\,k\,x\ \mathrm{dt} \;+\; \varepsilon\,\mathrm{d}W(t)\;. \qquad (4.181)$$

In the limit $\varepsilon \to 0$ the stochastic part disappears and the resulting ODE remains first order in time and we are dealing with a non-singular limit. The exact solution of (4.181) for the initial condition $x(0) = c_0$ is

$$x_\varepsilon(t) \;=\; c_0\,\exp\left(-k\,t\right) \;+\; \varepsilon\int_0^t \exp\Big(-k(t-\tau)\Big)\,\mathrm{d}W(\tau)\;. \qquad (4.182)$$

This case is particularly simple since the partitioning according to the series expansion (4.179b) is straightforward

$$x_0(t) \;=\; c_0\,\exp\left(-k\,t\right) \quad \text{and} \quad x_1(t) \;=\; \int_0^t \exp\Big(-k(t-\tau)\Big)\,\mathrm{d}W(\tau)\;,$$

and $x_0(t)$ is indeed the solution of the ODE obtained by setting $\varepsilon = 0$ in the SDE (4.181).

Now we consider the corresponding Fokker-Planck equation

$$\frac{\partial P(x,t)}{\partial t} \;=\; \frac{\partial}{\partial x}\Big(k\,x\,P(x,t)\Big) \;+\; \frac{1}{2}\,\varepsilon^2\,\frac{\partial^2 P(x,t)}{\partial t^2}\;, \qquad (4.183)$$

with the exact solution being a Gaussian with $x_0(t)$ as expectation value

$$\begin{aligned} E\big(x(t)\big) &= \alpha(t) = c_0 \exp(-k\,t) \text{ and} \\ \mathrm{var}\big(x(t)\big) &= \varepsilon^2\,\beta(t) = \varepsilon^2\,\frac{1-\exp(-2k\,t)}{2k}\ , \end{aligned} \tag{4.184}$$

and hence

$$P_\varepsilon(x,t|\,c_0,0) = \frac{1}{\varepsilon}\,\frac{1}{\sqrt{2\pi\beta(t)}}\,\exp\left(\frac{1}{\varepsilon^2}\,\frac{\big(x-\alpha(t)\big)^2}{2\beta(t)}\right)\ . \tag{4.184'}$$

In the limit $\varepsilon \to 0$ we obtain again the deterministic solution:

$$\lim_{\varepsilon\to 0} P_\varepsilon(x,t|\,c_0,0) = \delta\big(x-\alpha(t)\big)\ ,$$

which is the first order solution of the corresponding SDE and a deterministic trajectory along the path $x(t) = c_0 \exp(-kt)$. In the limit $\varepsilon \to 0$ the second order differential equation (4.183) is reduced to a first order equation, this implies a singularity and singular perturbation theory has to be applied. The probability density, however, cannot be expanded straightforwardly in a power series in ε, and the introduction of a scaled variable is needed before:

$$\xi = \big(x - \alpha(t)\big) \,/\, \varepsilon \text{ or } x = \alpha(t) + \varepsilon\,\xi\ .$$

Now we can write down the probability density in ξ up to second order,

$$P_\varepsilon(\xi,t|0,0) = P_\varepsilon(x,t|c_0,0)\cdot\frac{\mathrm{d}x}{\mathrm{d}\xi} = \frac{1}{\sqrt{2\pi\beta(t)}}\,\exp\left(-\frac{\xi^2}{2\beta(t)}\right)\ .$$

Scaling has eliminated the singularity as the probability density for ξ does not contain ε: The distribution of the scaled variable ξ is a Gaussian with mean zero and variance $\beta(t)$. The standard deviation from the deterministic trajectory $\alpha(t)$ is of order ε as ε goes to zero. The coefficient of ε is the random variable ξ. As it should be there is no difference in the interpretation between the Fokker-Planck and the stochastic differential equation.

4.5.3 Size expansion of the master equation

Although quite a few representative examples and model systems can be analyzed by solving one step birth-and-death master equations exactly (section 4.3), the actual applicability of this technique to specific problems of chemical kinetics is rather limited. In order to apply a chemical master equation to a problem in practice one is commonly dealing with at least 10^{12} particles. Upscaling discloses one particular issue that of size expansions that becomes obvious in the transition from master equations to Fokker-Planck

equations. The sample volume V is the best estimator of system size in condensed matter. Two classes of quantities are properly distinguished:

(i) *intensive properties* that are independent of system size, and
(ii) *extensive properties* that grow proportional to system size.

Examples of intensive properties are temperature, pressure, density or concentrations, whereas volume, particle numbers, energy or entropy are extensive properties. In upscaling from say 1000 to 10^{12} particles extensive properties grow by a factor of 10^9 whereas intensive properties remain the same. Some *pairs* of properties – one extensive and one intensive – are of particular importance, for example particle number $\mathcal{X}$ or n and concentration $a = \mathcal{X}_{\mathsf{A}}/(V \cdot N_L)$ or mass M and (volume) density $\varrho = M/V$, respectively. The system size that is used for scaling will be denoted by Ω, and if not stated otherwise we shall assume $\Omega = V \cdot N_L$. Properties describing the evolution of the system are modelled by variables and again we distinguish *extensive* and intensive variables. In case of the amount of a chemical compound, $[\mathbf{A}]$, we have the particle number $n(t) \propto \Omega$ as the extensive variable and the concentration $a(t) = n(t)/\Omega$ as the intensive variable, and we indicate this correspondence by $n \mathrel{\widehat{=}} a$.[61] The system size Ω itself is, of course, also an extensive property, the special extensive property, which has be chosen as reference.

Approximation methods have been developed, which turned out to be particularly illustrative and useful in the limit of sufficiently large systems. The Dutch theoretical physicist Nicholas van Kampen [439, 441] expands the master equation in the inverse square root of system size Ω. A discrete random variable $\mathcal{X}_{\mathsf{A}}$ with the probability density $P_n(t) = P\big(\mathcal{X}_{\mathsf{A}}(t) = n(t)\big)$ is considered in the limit to macroscopic description. The limit of interest is a large value of Ω at fixed a, which is tantamount to the transition to a macroscopic system.

The transition probabilities are reformulated as

$$W(n|m) \;\equiv\; \omega(\frac{m}{\Omega};\Delta n) \;\text{ with }\; \Delta n = n - m \;,$$

and scaled according to the assumption

$$W(n|m) \;=\; \Omega\,\omega\big(\frac{m}{\Omega};\Delta n\big) \;=\; \Omega\,\omega\big(a;\Delta n\big) \;. \tag{4.185}$$

The essential trick in the van Kampen expansion is that the size of the jump is expressed in term of an extensive quantity, Δn, whereas the intensive variable

[61] In order to improve clearness in the derivation of the size expansion we shall use the lowercase letters $a, b, c, \ldots$ for intensive variables and the lowercase letters $n, m, p, \ldots$ for extensive variables. When dealing with atoms, molecules or compounds intensive variables will continuous and mostly concentrations whereas the extensive variables are understood as particle numbers. In order to avoid misunderstanding we introduce the symbol '$\mathrel{\widehat{=}}$' to express the relation between conjugate intensive and extensive variables, for example $\varrho \mathrel{\widehat{=}} M$.

$a = n/\Omega$ is used for the calculation of the evolution of the system, $a(t)$. The expansion is made now in a new variable z, which is defined by

$$a \;=\; \Omega\,\phi(t) \,+\, \Omega^{1/2}\,z \;\;\text{or}\;\; z \;=\; \Omega^{-1/2}\,a \,-\, \Omega^{1/2}\,\phi(t)\;. \tag{4.186}$$

where the function $\phi(t)$ is still to be determined. The change of variables transforms the probability density $\Pi(a,t)$ and its derivatives according to

$$\begin{aligned}
&\Pi(a,t) \;=\; \Pi\big(\Omega\phi(t) + \Omega^{1/2}z, t\big) \;=\; P(z,t)\,,\\
&\frac{\partial^n P(z,t)}{\partial z^n} \;=\; \Omega^{n/2}\,\frac{\partial^n \Pi(a,t)}{\partial a^n}\,,\;\;\text{and}\\
&\frac{\partial P(z,t)}{\partial t} \;=\; \frac{\partial \Pi(a,t)}{\partial t} + \Omega\,\frac{\mathrm{d}\phi(t)}{\mathrm{dt}}\frac{\partial \Pi(a,t)}{\partial a} \;=\; \frac{\partial \Pi(a,t)}{\partial t} + \Omega^{1/2}\,\frac{\mathrm{d}\phi(t)}{\mathrm{dt}}\frac{\partial P(z,t)}{\partial z}\,.
\end{aligned}$$

The derivative moments $\alpha_n(a)$ are now proportional to the system size Ω and therefore we scale them accordingly: $\alpha_n(a) = \Omega\,\widetilde{\alpha}_n(x)$. In the next step the new variable z is introduced into the Kramers-Moyal expansion (4.177):

$$\begin{aligned}
\frac{\partial \Pi(a,t)}{\partial t} \;&=\; \frac{\partial P(z,t)}{\partial t} \,-\, \Omega^{1/2}\,\frac{\mathrm{d}\phi(t)}{\mathrm{dt}}\,\frac{\partial P(z,t)}{\partial z} \;=\\
&=\; \sum_{n=1}^{\infty} (-1)^n\,\frac{\Omega^{\,1-n/2}}{n!}\,\frac{\partial^n}{\partial z^n}\Big(\widetilde{\alpha}_n\big(\phi(t) + \Omega^{-1/2}\,z\big)\,P(z,t)\Big)\;,\\
\frac{\partial P(z,t)}{\partial t} \;&=\; \Omega^{1/2}\cdot\left(\frac{\mathrm{d}\phi(t)}{\mathrm{dt}} \,-\, \widetilde{\alpha}_1\big(\phi(t)\big)\right)\,\frac{\partial P(z,t)}{\partial z} \,+\, \Omega^{\,0}\cdot\Big(\cdots\Big)\cdots\;.
\end{aligned}$$

For general validity of an expansion all terms of a certain order in the expansion parameter must vanish. We make use of this property to define $\phi(t)$ such that the terms of order $\Omega^{1/2}$ are eliminated by demanding

$$\frac{\mathrm{d}\phi(t)}{\mathrm{dt}} \;=\; \widetilde{\alpha}_1\big(\phi(t)\big)\;. \tag{4.187}$$

This equation is an ODE determining $\phi(t)$ and, of course, it is in full agreement with the deterministic equation for the expectation value of the random variable and thus $\phi(t)$ is indeed the deterministic part of the solution.[62]

The next step is an expansion of $\widetilde{\alpha}_n\big(\phi(t) + \Omega^{-1/2}z\big)$ in $\Omega^{-1/2}$ and reordering of terms yielding

$$\frac{\partial P(z,t)}{\partial t} \;=\; \sum_{m=2}^{\infty}\frac{\Omega^{-(m-2)/2}}{m!}\sum_{n=1}^{m}(-1)^n\binom{m}{n}\,\widetilde{\alpha}_n^{m-n}\big(\phi(t)\big)\,\frac{\partial^n}{\partial z^n}\Big(z^{m-n}P(z,t)\Big)$$

[62] As shown in equations (3.94) and (3.102) this result is only true for linear first jump moments or for the linear approximation to the first jump moments (see below).

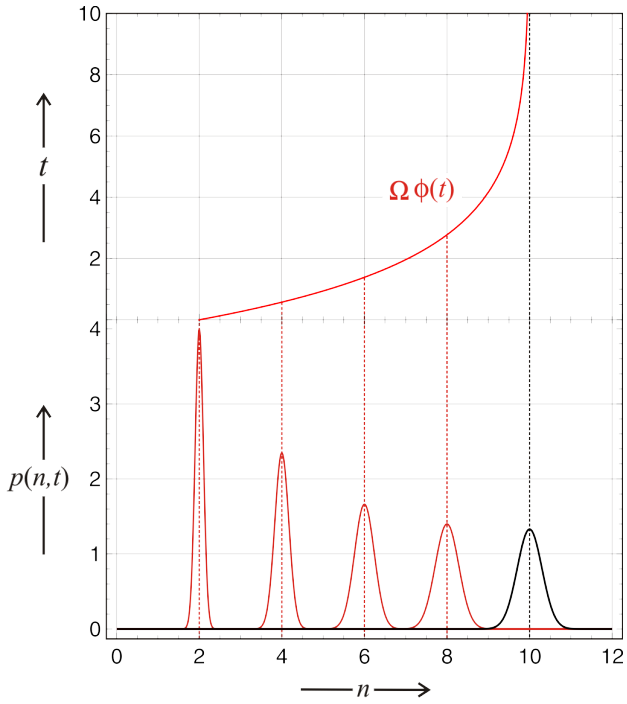


Fig. 4.37 The size expansion of a stochastic variable $\mathcal{X}(t)$. The variable n is partitioned according to into a macroscopic part and the fluctuations around it, $n(t) = \Omega\phi(t) + \Omega^{1/2}x(t)$, wherein Ω is a size parameter, for example the size of the population or the volume of the system (Equilibrium fluctuations are shown in black). Computations: $\Omega\phi(t) = 5n_0(1 - 0.8e^{-kt})$ with $n_0 = 2$ and $k = 0.5$; $p\,(n,t) = \Omega^{1/2}x(t) = e^{-(n-\Omega\phi(t))^2/(2\sigma^2)}/\sqrt{2\pi\sigma^2}$ with $\sigma = 0.1, 0.17, 0.24, 0.285, 0.30$.

In taking the limit of large system size Ω all terms vanish except the one with $m = 2$ and we find the result

$$\frac{\partial P(z,t)}{\partial t} \;=\; -\,\widetilde{\alpha}_1^{(1)}\big(\phi(t)\big)\,\frac{\partial}{\partial z}\Big(z\,P(z,t)\Big) \;+\; \frac{1}{2}\,\widetilde{\alpha}_2\big(\phi(t)\big)\,\frac{\partial^2}{\partial z^2}P(z,t)\;, \quad (4.188)$$

where $\alpha_1^{(1)}$ stands for the linear part of the drift term. In figure 4.37 we show a specific example of partitioning a process $n(t)$ into a macroscopic part $\Omega\phi(t)$ and fluctuations $\Omega^{1/2}x(t)$ around it.

It is straightforward to compare with the result of the Kramers-Moyal expansion (4.177) truncated after two terms:

$$\frac{\partial P(x,t)}{\partial t} \;=\; -\,\frac{\partial}{\partial x}\Big(\alpha_1(x)P(x,t)\Big) \;+\; \frac{1}{2}\,\frac{\partial^2}{\partial x^2}\Big(\alpha_2(x)P(x,t)\Big)\;.$$

The change of variables $\xi = x/\Omega$ leads to

$$\frac{\partial P(\xi,t)}{\partial t} \;=\; -\,\frac{\partial}{\partial \xi}\Big(\widetilde{\alpha}_{\,1}(\xi)P(\xi,t)\Big) \;+\; \frac{1}{2\Omega}\,\frac{\partial^2}{\partial \xi^2}\Big(\widetilde{\alpha}_{\,2}(\xi)P(\xi,t)\Big)\;.$$

Through application of small noise theory (section 4.5.2) with $\epsilon^2 = \Omega^{-1}$ and using the substitution $\xi = \Omega^{1/2}\big(x-\phi(t)\big)$ one obtains the lowest order Fokker-Planck equation, which is exactly the same as the lowest order approximation in the van Kampen expansion. This result has an important consequence: If we are only interested in the lowest order approximation we may use the Kramers-Moyal equation, which is much easier to derive than the van Kampen equation.

Eventually, we found a procedure to relate master equations and Fokker-Planck equations in an approximation that closes the gap between microscopic stochasticity and macroscopic behavior. It should be stressed, however, that the range of validity of a Fokker-Planck equation derived from a master equation is not independent of the kind of limiting procedure applied. If the transition is made by means of rigorous equations in a legitimate limit to continuous variables (section 4.5.4), the full nonlinear dependence of $\alpha_1(x)$ and $\alpha_2(x)$ can be seriously analyzed. If, on the other hand, an approximately valid approximation like the small noise approximation is applied it is appropriate to consider only the linearization of the drift term and individual solutions of this equations are represented by the trajectories of the stochastic equation:

$$\mathrm{d}z \;=\; \widetilde{\alpha}_1^{(1)}\big(\phi(t)\big)\,z\;\mathrm{dt} \;+\; \sqrt{\widetilde{\alpha}_{\,2}\big(\phi(t)\big)}\;\mathrm{d}W(t)\;. \tag{4.189}$$

The choice of the best way of scaling also depends on the special case to be studied and we close this section by presenting a few examples: (i) the flow reactor,(ii) the reversible first order chemical reaction, and (iii) the birth and death process in epidemiology.

Equilibration in the flow reactor. The problem we are reconsidering here is the time dependence of a single chemical substance **A** in a device for performing chemical reactions under controlled conditions as described in section 4.3.1. The concentration of **A** in the reactor, $a(t)$, starts from some initial value $a_0 = a(0)$ and, after flow equilibrium has been established, $\lim_{t\to\infty} a(t) = \bar{a}$, it adopts the value $\bar{a} = \widehat{a}$, where $\widehat{a}$ is the concentration of **A** in the stock solution flowing into the reactor (figure 4.18). The flow in and out of the reactor is controlled by the flow rate r commonly measured in volume·time^{-1} = [cm^3/sec] and it represents the reciprocal mean residence time of the solution in the reactor: $\tau_{\mathrm{v}}^{-1} = r/V$ with V being the total volume of the reactor.

The number of particles **A** in the reactor is a stochastic variable, $\mathcal{N}_{\mathsf{A}}(t)$, with the probability density $P_n(t) \;=\; P\big(\mathcal{N}_{\mathsf{A}}(t) = n\big)$. It is – at the same time – the discrete extensive variable: $n_{\mathbf{A}}(t) = n(t)$ with $n \in \mathbb{N}$. The concentration is the continuous intensive variable: $a(t) = n(t)/V \cdot N_{\mathrm{L}}$ with $\Omega = V \cdot N_{\mathrm{L}}$. The equilibration of the reactor can be described by the master equation

$$\frac{\partial P_n(t)}{\partial t} = W(n|n-1)\,P_{n-1}(t) + W(n|n+1)\,P_{n+1}(t) - \\ - \big(W(n-1|n) + W(n+1|n)\big)\,P_n(t)\,;\ n \in \mathrm{N}\ , \qquad (4.87')$$

with the elements of the tridiagonal transition matrix W given by

$$W(n|m) = r\left(\delta_{n,m+1}\,\widehat{n} + \delta_{n,m-1}\,n\right)\ . \qquad (4.190)$$

The only nonzero contribution of the first term requires $n = m + 1$ and describes an increase by one of the particle number in the reactor through inflow that corresponds to the step-up transition probability $w_n^+ = r\widehat{n}$. The nonzero contribution of the second term, $n = m - 1$, deals with the loss of a particle **A** through outflow in the sense of a step-down transition with the probability $w_n^- = rn$. The equilibration of the flow reactor thus can be understood as a linear death process with *immigration* expressed by a positive constant term, $r\widehat{n}$.

The reformulation of the transition matrix (4.185) in the sense of van Kampen's expansion leads to

$$W(a;\Delta n) = \Omega\left(r\,\widehat{a}\,\delta_{\Delta n,+1} + r\,a\,\delta_{\Delta n,-1}\right)$$

with $\Delta n = n - m$. Calculation of the first two jump moments yields

$$\alpha_1(n) = \sum_{m=0}^{\infty}(m-n)\,W(m|n) = r(\widehat{n}-n) = \Omega\,r(\widehat{a}-a)\ ,$$

$$\alpha_2(n) = \sum_{m=0}^{\infty}(m-n)^2\,W(m|n) = r(\widehat{n}+n) = \Omega\,r(\widehat{a}+a)\ ,$$

and the deterministic equation with $\phi(t) = a(t) = n(t)/\Omega$ is of the form

$$\frac{\mathrm{d}a}{\mathrm{dt}} = r\,(\widehat{a} - a) \ \text{ and } \ a(t) = \widehat{a} + \big(a(0) - \widehat{a}\big)\,e^{-rt}\ ,$$

where we recall that the equilibrium concentration of **A** in the reactor is equal to the influx concentration: $\bar{a} = \widehat{a}$. Following the procedure of van Kampen's expansion we define

$$n = \Omega\,\phi(t) + \Omega^{1/2}\,z \ \text{ or } \ z = \Omega^{-1/2}n - \Omega^{1/2}\,\phi(t) \qquad (4.186')$$

and obtain the Fokker-Planck equation

$$\frac{\partial P(z,t)}{\partial t} = r\,\frac{\partial}{\partial z}\Big(z\,P(z,t)\Big) + \frac{r}{2}\,\frac{\partial^2}{\partial z^2}\Big(\big(\widehat{a} + a(t)\big)\,P(z,t)\Big)\ , \qquad (4.191)$$

which leads to the expectation value and variance in the scaled variable z:

$$\begin{aligned} E\big(z(t)\big) &= z(0)\, e^{-rt} \;, \\ \mathrm{var}\big(z(t)\big) &= \big(\widehat{a} + a(0)\, e^{-rt}\big)\,(1 - e^{-rt}) \;. \end{aligned}$$

Since the partition of the variable n in equation (4.186') is arbitrary we can assume $z(0) = 0$[63] Transformation into the extensive variable, the particle number n yields

$$\begin{aligned} E\big(n(t)\big) &= \widehat{n} + \big(n(0) - \widehat{n}\big)\, e^{-rt} \;, \\ \mathrm{var}\big(n(t)\big) &= \big(\widehat{n} + n(0)\, e^{-rt}\big)\,(1 - e^{-rt}) \;. \end{aligned} \tag{4.192}$$

The stationary solution of the Fokker-Planck equation is readily calculated

$$\bar{P}(z) = \frac{1}{\sqrt{2\pi\,\widehat{a}}} \exp\left(-\frac{z^2}{2\widehat{a}}\right)$$

and it represents the approximation of the exact stationary Poisson density by means of a Gaussian as mentioned in (2.40):

$$\bar{P}(n) = \frac{\widehat{n}^n}{n!} \exp\left(-\widehat{n}\right) \approx \frac{1}{\sqrt{2\pi\,\widehat{n}}} \exp\left(-\frac{(n-\widehat{n})^2}{2\widehat{n}}\right).$$

A comparison of the different expansion techniques is made in the next paragraph where we consider the simple chemical reaction $\mathbf{A} \rightleftharpoons \mathbf{B}$ with compound $\mathbf{B}$ buffered that gives rise to a master equation that is formally identical to that for the equilibration of the flow reactor.

The chemical reaction $\mathbf{A} \rightleftharpoons \mathbf{B}$. The reversible monomolecular conversion reaction is considered under large excess of compound $\mathbf{B}$: The concentration $[\mathbf{B}] = b_0 = n_\mathsf{B}/\Omega$ is constant or in a condition named *buffered*. The stochastic variable counts the number of molecules $\mathbf{A}$ in the system: $[\mathbf{A}] = \mathcal{N}_\mathsf{A}(t)$ with the probability distribution $P_{n_\mathsf{A}}(t) = P_n(t) = P\big(\mathcal{N}_\mathsf{A}(t) = n\big)$ and $a(t) = n(t)/\Omega$. The elements of transition matrix of the master equation (4.87') are:

$$W(n|m) = \delta_{n,m+1}\, l\, n_\mathsf{B} + \delta_{n,m-1}\, k\, n \;, \tag{4.190'}$$

where k and l are the rate parameters for the forward and backward reaction, respectively. By replacing the constant terms $l n_\mathsf{B} \Leftrightarrow r\widehat{n}$ and $k \Leftrightarrow r$ we recognize that the two problems, flow reactor and buffers reaction $\mathbf{A} \rightleftharpoons \mathbf{B}$, are formally identical. By application of van Kampen's expansion the solutions are derived in precisely the same way as in the previous paragraph. With $n = \Omega\phi(t) + \Omega^{1/2} z$ we obtain for the deterministic solution and the Fokker-Planck equation,

[63] The assumption $z(0) = 0$ implies $z(t) = 0$ and hence the corresponding stochastic variable $\mathcal{Z}(t)$ describes the fluctuations around zero.

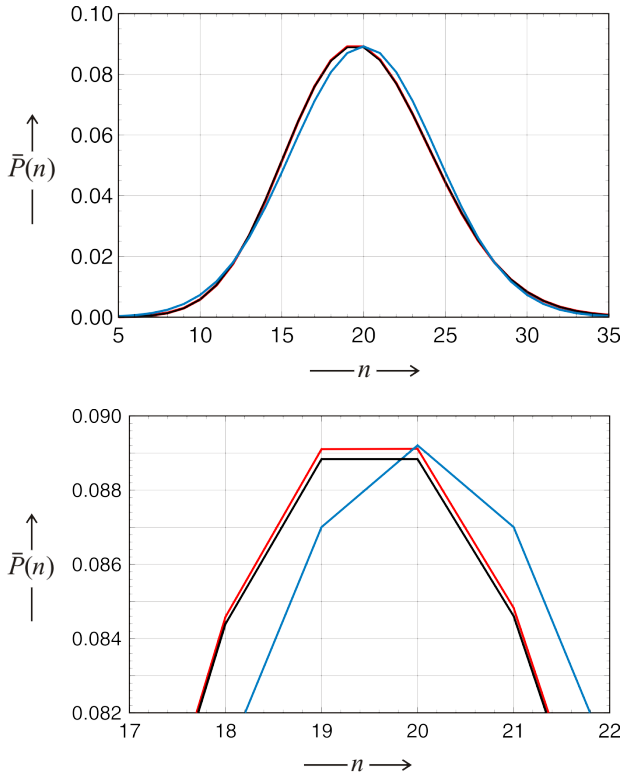


Fig. 4.38 Comparison of expansions of the master equation. The reaction $\mathbf{A} \rightleftharpoons \mathbf{B}$ with compound **B** buffered, $[\mathbf{B}] = b = b_0 = n_\mathsf{B}/\Omega$, is chosen as example and the exact stationary solution (black) is compared with the results of the Kramers-Moyal expansion (red) and the van Kampen size expansion (blue). Parameter choice: $V = 1$, $k_1 = 2$, $l_1 = 1$, $n_\mathsf{B} = 40$.

$$\frac{\mathrm{d}\phi(t)}{\mathrm{dt}} \;=\; l b_0 \,-\, k\,\phi(t) \quad \text{and} \quad \phi(t) \;=\; \phi(0)\,e^{-k\,t} \,+\, \frac{l b_0}{k}\,(1 - e^{-k\,t}) \;.$$

$$\frac{\partial P(z)}{\partial t} \;=\; k\,\frac{\partial}{\partial z}\Big(z\,P(z)\Big) \,+\, \frac{1}{2}\,\frac{\partial^2}{\partial z^2}\Big(\big(l b_0 \,+\, k\,\phi(t)\big)\,P(z)\Big) \;,$$

respectively.

The expectation value of z is $\mathrm{E}\big(z(t)\big) \;=\; z(0)e^{-k\,t}$. It vanishes with the usual assumption $z(0) = 0$, for the variance $\mathrm{var}\big(z(t)\big)$ we find

$$\mathrm{var}\big(z(t)\big) \;=\; \left(\frac{l b_0}{k} \,+\, \phi(0)\right)\,(1 - e^{-kt}) \;,$$

and for the solutions in the variable n with $n(0) = \Omega\,\phi(0)$ we obtain

$$\begin{aligned} \mathrm{E}\big(n(t)\big) &= \Omega\,\phi(t) = n(0)\,e^{-kt} + \frac{l_1\,n_{\mathsf{B}}}{k}\,(1 - e^{-kt})\;, \\ \mathrm{var}\big(n(t)\big) &= \Omega\,\mathrm{var}\big(z(t)\big) = \left(\frac{l_1\,n_{\mathsf{B}}}{k} + n(0)\right)(1 - e^{-kt})\;. \end{aligned}$$

Finally, we compare the stationary state solutions obtained from the van Kampen expansion and from the Kramers-Moyal expansion with the exact solution. The size expansion yields

$$\bar{P}(z) = \frac{1}{\sqrt{\frac{\pi\kappa}{2}}\,\big(1 + \mathrm{erf}(\sqrt{\frac{\kappa}{2}})\big)}\,\exp\left(-\frac{(n-\kappa)^2}{2\kappa}\right)\;, \tag{4.193a}$$

where we use $\kappa = ln_{\mathsf{B}}/k$ and replaced $z \Leftrightarrow n$. The result of the truncated Kramers-Moyal expansion is calculated from the stationary solution (3.82) of a Fokker-Planck equation with $A(n) = \alpha_1(n) = l\,n_{\mathsf{B}} - k\,n$ and $B(n) = \alpha_2 = l\,n_{\mathsf{B}} + k\,n$. with those derived from the Kramers-Moyal expansion

$$\bar{P}(n) = N \cdot (l\,n_{\mathsf{B}} + k\,n)^{-1+4\,l\,n_{\mathsf{B}}/k}\,e^{-2n}\;, \tag{4.193b}$$

where the normalization factor N is still to be determined for the special case. The exact solution is identical with the result derived for the flow reactor (4.91),

$$\bar{P}(n) = \frac{\big(l\,n_{\mathsf{B}}/k\big)^n\,\exp\big(-l\,n_{\mathsf{B}}/k\big)}{n!} = \frac{\kappa^n\,e^{-\kappa}}{n!}\;, \tag{4.193c}$$

which is a Poissonian. A comparison of numerical plots is shown in figure 4.38. It is remarkable how well the truncated Kramers-Moyal expansion agrees with the exact probability density. It is easy to understand therefore that it is much more popular than the size expansion, which is much more sophisticated. We remark that the major difference between the van Kampen solution and the other two curves results in essence from the approximation of a Poissonian by a Gaussian (see figure 2.8).

4.5.4 From master to Fokker-Planck equations

Finally, we summarize this section by mentioning another general scaling method [157, pp. 273-274] that reminds of the transition from continuous time random walks to diffusion, which has been discussed in section 3.2.4: The master equation (3.104) is converted to the partial differential equation of the Wiener process equation (3.55), a Fokker-Planck equation without drift by taking the limit of infinitesimally small steps at infinite frequency, and which is formally identical with the 1D diffusion equation. In this transition the step size was chosen to be $l = l_0 \cdot \varepsilon$ and the probability to make a step was $\vartheta = \vartheta_0/\varepsilon^2$. During the transition the jumps become simultaneously smaller and more probable and both changes are taken care by a *scaling assumption*, which is based on the usage of a scaling parameter ε: The average step size is proportional to ε and so is the variance of the step size,[64] and thus decreases with ε whereas the jump probabilities increase as ε becomes smaller.

Here we perform the transition from master equations to Fokker-Planck equations in a more general way and illustrate by means of examples that a diffusion process can always be approximated by a master equation whereas the reverse is not true. First the elements of the transition matrix are rewritten in terms of a new variable $\eta = (z - x - A(x)\,\varepsilon)/\sqrt{\varepsilon}$, where $A(x)$ represents the general drift term. The transition probabilities we written in the form

$$W_\varepsilon(z\,|x) \;=\; \varepsilon^{-3/2}\,\phi(\eta, x) \tag{4.194}$$

where the function $\phi(\eta, x)$ is given by the concrete example to be studied and, in addition, fulfils the relations

$$\int \mathrm{d}\eta\; \phi(\eta, x) \;=\; I \quad \text{and} \quad \int \mathrm{d}\eta\; \eta\; \phi(\eta, x) \;=\; 0\;.$$

We define consistent expressions for the first three jump moments (4.176),

$$\alpha_0(x) \;\doteq\; \int \mathrm{d}z\, W_\varepsilon(z|x) \;=\; \frac{I}{\varepsilon} \tag{4.195a}$$

$$\alpha_1(x) \;\doteq\; \int \mathrm{d}z\,(z - x)\, W_\varepsilon(z|x) \;=\; A(x)\, I \tag{4.195b}$$

$$\alpha_2(x) \;\doteq\; \int \mathrm{d}z\,(z - x)^2\, W_\varepsilon(z|x) \;=\; \int \mathrm{d}\eta\, \eta^2\, \phi(\eta, x)\;. \tag{4.195c}$$

These expressions are obtained from the definitions of the variable η and the two integrals of $\phi(\eta, x)$ and, in case of (4.195c) through neglect of the term of order $\mathcal{O}(\varepsilon) = A(x)^2 I\,\varepsilon$ in the limit $\varepsilon \to 0$. For taking this limit we shall assume further that the function $\phi(\eta, x)$ vanishes sufficiently fast as $y \to \infty$ in order to guarantee that

[64] This is automatically fulfilled when the steps follow a Poisson distribution.

$$\lim_{\varepsilon\to 0} W_\varepsilon(z|x) \;=\; \lim_{\eta\to\infty}\left(\left(\frac{x}{z-x}\right)^3 \phi(\eta,x)\right) \;=\; 0 \quad \text{for} \quad z\neq x\;.$$

Very similar to the derivation of the differential Chapman-Kolmogorov equation in section 3.2 we may choose some twice differentiable function $f(z)$ and show that

$$\lim_{\varepsilon\to 0}\left\langle \frac{\partial f(z)}{\partial t}\right\rangle \;=\; \left\langle \alpha_1(z)\,\frac{\partial f(z)}{\partial z} \;+\; \frac{1}{2}\,\alpha_2(z)\,\frac{\partial^2 f(z)}{\partial z^2}\right\rangle\;.$$

Applying this result to the probability $P(x,t)$ result has the consequence that in the limit $\varepsilon\to 0$ the master equation

$$\frac{\partial P(x,t)}{\partial t} \;=\; \int \mathrm{d}z\,\Big(W(x|z)\,P(z,t) \;-\; W(z|x)\,P(x,t)\Big) \tag{4.196a}$$

becomes the Fokker-Planck equation

$$\frac{\partial P(x,t)}{\partial t} \;=\; -\,\frac{\partial}{\partial x}\Big(\alpha_1\,P(x,t)\Big) \;+\; \frac{1}{2}\,\frac{\partial^2}{\partial x^2}\Big(\alpha_2\,P(x,t)\Big)\;. \tag{4.196b}$$

Accordingly, one can construct a Fokker-Planck limit for the master equation if and only if the requirements imposed by the three jump moments α_p, $p=0,1,2$ (4.195) can be met. In case these criteria are not fulfilled, there is no approximation possible as illustrated now by means of examples.

Continuous time random walk. The master equation introduced in subsection 3.2.4,

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; \vartheta\Big(P_{n+1}(t) \;+\; P_{n-1}(t) \;-\; 2\,P_n(t)\Big) \quad \text{with} \quad P_n(t_0) \;=\; \delta_{n,n_0}$$

as initial condition, is to be converted to a Fokker-Planck equation. First we remember that the steps were embedded in a continuous spatial coordinate, $x = n\cdot l$, and accordingly the walk started at the point for $x_0 = n_0\cdot l$. The elements of the transition matrix W are of the general form

$$W(z|x) \;=\; \vartheta\,(\delta_{z,x-l} + \delta_{z,x+l})\;,$$

and by means of the three integrals over the scaled transition moments

$$\int_{-\infty}^{\infty}\mathrm{d}\eta\,\phi(\eta,x) \;=\; \varepsilon\,2\vartheta\;,\quad \int_{-\infty}^{\infty}\mathrm{d}\eta\,\eta\,\phi(\eta,x) \;=\; (l-l)\,\vartheta \;=0\;,$$
$$\int_{-\infty}^{\infty}\mathrm{d}\eta\,\eta^2\,\phi(\eta,x) \;=\; 2\,l^2\,\vartheta\;,$$

where the second integral vanishes because of the intrinsic symmetry of the random walk. The first three jump moments are readily calculated from equa-

tion (4.195)

$$\alpha_0(x) = 2\vartheta, \quad \alpha_1(x) = 0, \quad \alpha_2(x) = 2\,l^2\,\vartheta\,.$$

With the introduction of the variable η we got a natural way of scaling step size and jump probability: Assume we began with some some discrete system (l_0, ϑ_0), reduce the step size according to $l^2 = \varepsilon \cdot l_0^2$ and raise the probability by $\vartheta = \vartheta_0/\varepsilon$. Then, the diffusion coefficient $D = (l_0^2 \cdot \varepsilon) \cdot (\vartheta_0/\varepsilon$ remains constant in the scaling process. With $D = l^2 \cdot \vartheta = l_0^2 \cdot \vartheta_0$ we obtain a Fokker-Planck equation, the familiar stochastic diffusion equation

$$\frac{\partial P(x,t)}{\partial t} \;=\; D\,\frac{\partial^2 P(x,t)}{\partial x^2}\;. \tag{3.55'}$$

The final result obtained is exactly the same as in section 3.2.4, although there we used a much simpler intuitive procedure instead of the transformation (4.194).

Poisson process. The Poisson process can be viewed as a random walk restricted to one direction and therefore taking place in a the (upper) half-plane with the master equation

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; \vartheta\Big(P_{n+1}(t) \,-\, P_n(t)\Big) \;\text{ with }\; P_n(t_0) \,=\, \delta_{n,n_0}$$

The notation used in section 3.2.2.4 is slightly modified: $\alpha \Leftrightarrow \vartheta$, and with $x = n \cdot l$ we find for the transition matrix W:

$$W(x|z) \;=\; \vartheta\,\delta_{z,x+l}\;.$$

The calculation of the moments is exactly the same as in case of the previous example:

$$\alpha_0(x) = \vartheta, \quad \alpha_1(x) = l\,\vartheta, \quad \alpha_2(x) = l^2\,\vartheta\,.$$

In this case there is no way to define l and ϑ as functions of ε such that both $\alpha_1(x)$ and $\alpha_2(x)$ remain finite in the limit $l \to 0$. Applying, for example, the same model assumption as made for the one-dimensional random walk we find $l^2 = l_0^2\,\varepsilon$ and $\vartheta = \vartheta_0/\varepsilon$, and hence $\lim_{\varepsilon\to 0} l^2 \cdot \vartheta = D$ as before but $\lim_{\varepsilon\to 0} l\vartheta = \lim_{\varepsilon\to 0} l_0 \cdot \vartheta_0/\sqrt{\varepsilon} = \infty$. Accordingly, there is no Fokker-Planck limit for the Poisson process within the transition moment expansion scheme.

General birth-and-death master equations. Crispin Gardiner provides also an scaling analysis leading to the general Fokker-Planck equation. The starting point is a master equation with the transition probability matrix

$$W_\varepsilon(z|x) \;=\; \left(\frac{A(x)}{2\varepsilon} + \frac{B(x)}{2\varepsilon^2}\right)\,\delta_{z,x+\varepsilon} \;+\; \left(-\frac{A(x)}{2\varepsilon} + \frac{B(x)}{2\varepsilon^2}\right)\,\delta_{z,x-\varepsilon}\;, \tag{4.197}$$

where $W_\varepsilon(z|x)$ is positive at least for sufficiently small ε: $W_\varepsilon(z|x) > 0$ if $B(x) > \varepsilon\,|A(x)|$. Under the assumption that this is fulfilled for the entire do-

main of the variable x, the process takes place on an x-axis that is partitioned into integer multiples of ε.[65] In the limit $\varepsilon \to 0$ the birth-and-death master equation is converted into a Fokker-Planck equation with

$$\begin{aligned} &\alpha_0(x) = B(x)/\varepsilon^2, \quad \alpha_1(x) = A(x), \quad \alpha_2(x) = B(x) \ \text{ and} \\ &\lim_{\varepsilon\to 0} W_\varepsilon(z\,|x) = 0 \text{ for } z \neq x\,. \end{aligned} \tag{4.198}$$

Nevertheless, the imagination of jumps converging smoothly into a continuous distribution is no longer valid, because the zeroth moment $\alpha_0(x)$ diverges with $1/\varepsilon^2$ and not with $1/\varepsilon$ as required by equation (4.195a). Notwithstanding there exists a limiting Fokker-Planck equation, because the limiting behavior of $\alpha_0(x)$ has no influence since it does not show up in the final equation

$$\frac{\partial P(x,t)}{\partial t} \;=\; -\,\frac{\partial}{\partial x}\Big(A(x)\,P(x,t)\Big) \;+\; \frac{1}{2}\,\frac{\partial^2}{\partial x^2}\Big(B(x)\,P(x,t)\Big)\;. \tag{3.47'}$$

Equation (4.198) provides a tool for the simulation of a diffusion process by an approximating birth-and-death process. The method, however, fails for $B(x) = 0$ for all possible ranges of x since then $W_\varepsilon(z,x)$ cannot fulfil the criterion of being nonnegative. Otherwise there is no restriction on the side of the Fokker-Planck equation, since equation (4.197) is completely general. As said the converse is not true: There are jump processes and master equations, which cannot be approximated by Fokker-Planck equations through scaling. The Poisson process discussed above may serve as an example.

Summarizing this section we compare the size expansion described in section 4.5.3 and the moment expansion presented here: In the size expansion (4.188) system size Ω was considered as a parameter and lim $\Omega \to \infty$ has been the transition of interest that leads to the macroscopic or deterministic equations. In the moment expansion, equations (4.195) and (4.196b), the system size was assumed to be constant and the transition concerned the resolution of the jump size that was increased from coarse grained to smooth or continuous variables, lim $\varepsilon \to 0$.

[65] We remark that the scaling relations (4.194) and (4.197) not the same but both lead to a Fokker-Planck equation.

4.6 Numerical simulation of chemical master equations

Historically the basis for numerical simulation of master equations was laid by the works of Andrey Kolmogorov and Willy Feller: Kolmogorov [258] introduced the differential equation describing Markov jump processes and Feller [129] defined the conditions under which the solutions of the Kolmogorov equations fulfilled the conditions for proper probabilities. In addition, he was able to prove that the time between consecutive jumps is exponentially distributed and that the probability of the next event is proportional to the deterministic rate. In other words, he provided evidence that sampling of jump trajectories leads to a statistically correct representation of the stochastic process. Joe Doob extended Feller's derivation beyond the validity for pure jump processes [93, 94]. The implementation of a stochastic simulation algorithm for the Kolmogorov equations is due to David Kendall [241] and was applied to studies of epidemic outbreaks by Maurice Bartlett [31]. More than twenty years later, almost at the same time when the Feinberg-Horn-Jackson theory of chemical reaction networks was introduced, the American physicist and mathematical chemist Daniel Gillespie [166, 167, 169, 173] revived the formalism and introduced a popular simulation tool for stochastic chemical reactions. His algorithm became popular as a simple and powerful tool for the calculation of single trajectories. In addition he showed that the chemical master equation and the simulation algorithm can be put together on a firm physical and mathematical basis [169]. Meanwhile the Gillespie algorithm became an essential simulation tool in chemistry and biology. Here we present the concept and the implementation of the algorithm, and demonstrate the usefulness by means of selected examples.

4.6.1 Basic assumptions

Gillespie's general stochastic model is introduced here by means of the same definitions and notations as used in the theory of chemical reaction networks (section 4.1.3): A set of M different molecular species, $\boldsymbol{\Xi} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_M\}$ in a homogeneous medium are interconverted through K elementary chemical reactions $\mathcal{R} = \{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_K\}$. Two conditions are assumed to be fulfilled by the system: (i) The content of a container with constant volume V is thought to be *well mixed* and spatially homogeneous (CSTR in figure 4.18), and (ii) the system is assumed to be in *thermal equilibrium* at constant temperature T. The primary goals of the simulation are the computation of the time courses of the stochastic variables – $\mathcal{X}_k(t)$ counting the number of molecules $\mathbf{X}_k$ of species $\mathbf{K}$ at time t – and the description of the evolution of the entire molecular population. The computations yield *exact* trajectories of the type shown in figure 4.14 (section 4.2.1). Within the frame of the two conditions for choosing a proper time interval for τ-leaping (sections 4.2.3

and 4.6.2) the trajectories provide solutions that correspond to the proper stochastic differential equations.

Variables, reactions, and stoichiometry. The entire population of a reaction system involving M species in K reactions is described by an M-dimensional random vector counting the numbers of molecules of individual species $\mathbf{X}_k$,

$$\vec{\mathcal{X}}(t) \;=\; \big(\mathcal{X}_1(t), \mathcal{X}_2(t), \ldots, \mathcal{X}_k(t) \;\ldots\; \mathcal{X}_M(t)\big)\;.$$

Molecules are discrete quantities and the random variables are discrete in the calculation of *exact* trajectories as well as in the chemical master equation: $\mathbf{n} = \big(\mathrm{n}_1(t), \mathrm{n}_2(t), \ldots, \mathrm{n}_M(t)\big)$. Three quantities are required to fully characterize a particular reaction channel $\mathbf{R}_{\boldsymbol{\mu}}$: (i) the specific probabilistic rate parameter, $\gamma_{\boldsymbol{\mu}}$, (ii) the kinetic functions $h_{\boldsymbol{\mu}}(\mathbf{n})$, and (iii) the stoichiometric matrix S.

In section 4.1.4 we derived the fundamental fact that a scalar rate parameter $\gamma_{\boldsymbol{\mu}}$, which is independent of dt, exists for each elementary reaction channel $\mathbf{R}_{\boldsymbol{\mu}}$ with $\boldsymbol{\mu} = 1, \ldots, K$ that is accessible to the molecules of a well-mixed and thermally equilibrated system in gas phase or solution, such that

$$\begin{aligned}\gamma_{\boldsymbol{\mu}}\ \mathrm{dt}\ \;=\; &\text{probability that a randomly selected combination of}\\ &\mathbf{R}_{\boldsymbol{\mu}}\ \text{reactant molecules at time } t \text{ will react within}\\ &\text{the next infinitesimal time interval } [t, t+\mathrm{dt}[\,.\end{aligned} \tag{4.199}$$

In addition to $\gamma_{\boldsymbol{\mu}}$ we shall require a function $h_{\boldsymbol{\mu}}(\mathbf{n})$ where the vector $\mathbf{n}(t)$ contains the exact numbers of all molecules at time t,

$$\begin{aligned}h_{\boldsymbol{\mu}}(\mathbf{n})\ \equiv\ &\text{the number of distinct combinations of } \mathbf{R}_{\boldsymbol{\mu}}\ \text{reactant}\\ &\text{molecules in the system when the numbers of molecules}\\ &\text{of species } \mathbf{X}_k \text{ are exactly } \mathrm{n}_k \text{ with } k = 1, \ldots, M\ ,\end{aligned} \tag{4.200}$$

and the stoichiometric matrix $\mathrm{S} = \{s_{k\boldsymbol{\mu}};\, k = 1, \ldots, M,\, \boldsymbol{\mu} = 1, \ldots, K\}$, an $M \times K$ matrix of integers, where

$$\begin{aligned}s_{k\boldsymbol{\mu}}\ \equiv\ &\text{the change in the } \mathbf{X}_k \text{ molecular population caused by the}\\ &\text{occurrence of one } \mathbf{R}_{\boldsymbol{\mu}} \text{ reaction.}\end{aligned} \tag{4.201}$$

The functions $h_{\boldsymbol{\mu}}(\mathbf{n})$ and the matrix S are derived from the stoichiometric equations (4.5) of the individual reaction channels as shown in section 4.1 and illustrated here by means of an example:

$$\begin{aligned}&\mathbf{R}_1:\ \mathbf{X}_1 + \mathbf{X}_2 \longrightarrow \mathbf{X}_3 + \mathbf{X}_4\ ,\\ &\mathbf{R}_2:\ 2\,\mathbf{X}_1 \longrightarrow \mathbf{X}_1 + \mathbf{X}_5\ ,\quad \text{and}\\ &\mathbf{R}_3:\ \mathbf{X}_3 \longrightarrow \mathbf{X}_5\ .\end{aligned} \tag{4.202}$$

In particular we find for the functions $h_{\boldsymbol{\mu}}(\mathbf{n})$[66]

$$\begin{aligned} h_1(\mathbf{n}) &= n_1\, n_2 \ , \\ h_2(\mathbf{n}) &= n_1\,(n_1-1) \ , \quad \text{and} \\ h_3(\mathbf{n}) &= n_3 \ , \end{aligned}$$

and for the stoichiometric matrix S

$$\mathrm{S} = \begin{pmatrix} -1 & -1 & 0 \\ -1 & 0 & 0 \\ +1 & 0 & -1 \\ +1 & 0 & 0 \\ 0 & +1 & +1 \end{pmatrix} \ ,$$

where the rows refer to molecular species, $\boldsymbol{X} = (\mathbf{X}_1, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5)$, and the columns to individual reactions, $\mathcal{R} = (\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3)$. The product side is considered in the stoichiometric matrix S by a positive sign of the stoichiometric coefficients whereas reactants are accounted for by a negative sign. Sometimes we shall make use of vectors corresponding to individual reactions $\mathbf{R}_{\boldsymbol{\mu}}$: $\mathbf{s}_{\boldsymbol{\mu}} = (s_{1\boldsymbol{\mu}}, \ldots, s_{M\boldsymbol{\mu}})^{\mathrm{t}}$. It is worth noticing that the functional form of $h_{\boldsymbol{\mu}}$ is determined exclusive by the reactant side of $\mathbf{R}_{\boldsymbol{\mu}}$. For mass action kinetics there is only one difference between the deterministic and the stochastic expressions: Since the particles are counted exactly in the latter approach we have to use $n(n-1)$ instead of n^2 because $n-1$ is significantly different from n only in small systems.

It is illustrative to consider the relation to conventional deterministic chemical kinetics. If we denote the concentration vector of the molecular species $\boldsymbol{X}$ by $\mathbf{x} = (x_1, \ldots, x_M)^{\mathrm{t}}$ and the flux or rate vector by $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_M)^{\mathrm{t}}$ the kinetic equation can be expressed by

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathrm{S} \cdot \boldsymbol{\varphi} \ . \tag{4.203}$$

The individual elements of the flux vector in mass action kinetics are

$$\varphi_{\boldsymbol{\mu}} = k_{\boldsymbol{\mu}} \prod_{k=1}^{n} x_k^{\nu_{k\boldsymbol{\mu}}} = k_{\boldsymbol{\mu}}\, \mathbf{x}^{\boldsymbol{\nu}_{\boldsymbol{\mu}}} \quad \text{for}$$

$$\nu_{1\boldsymbol{\mu}}\, \mathbf{X}_1 + \nu_{2\boldsymbol{\mu}}\, \mathbf{X}_2 + \ldots + \nu_{M\boldsymbol{\mu}}\, \mathbf{X}_M \longrightarrow$$

[66] As mentioned before there are two ways for taking appropriate account of combinatorics: (i) $h(\mathbf{n}) = \prod_i \binom{n_i}{\nu_i}$ and $\gamma/\nu_i!$ as rate parameter or (ii) $h(\mathbf{n}) = \prod_i n_i!/(n_i-\nu_i)!$ and γ. We use here version (ii) unless stated otherwise.

wherein the factors $\nu_{k\boldsymbol{\mu}}$ are nonnegative integers, the stoichiometric coefficients on the reactant side of the stoichiometric equations.

Reactions events. The probability of occurrence of reaction events within an infinitesimal time interval dt fulfils three general conditions for master equations that were formulated and discussed in section 4.2.1. Here we repeat: *Condition 1:* If $\vec{\mathcal{X}}(t) = \mathbf{n}$, then the probability that *exactly one* $\mathbf{R}_{\boldsymbol{\mu}}$ will occur in the system within the time interval $[t, t + \mathrm{dt}[$ is equal to

$$\gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\mathbf{n})\ \mathrm{dt}\ +\ o(\,\mathrm{dt})\ ,$$

where $o(\,\mathrm{dt})$ denotes terms that approach zero with dt faster than dt. *Condition 2:* If $\vec{\mathcal{X}}(t) = \mathbf{n}$, then the probability that *no* reaction will occur within the time interval $[t, t + \mathrm{dt}[$ is equal to

$$1\ -\ \sum_{\boldsymbol{\mu}} \gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\mathbf{n})\ \mathrm{dt}\ +\ o(\,\mathrm{dt})\ .$$

Condition 3: The probability of more than one reaction occurring in the system within the time interval $[t, t\ +\ \mathrm{dt}[$ is of order $o(\,\mathrm{dt})$. We express the probability $P(\mathbf{n}, t\ +\ \mathrm{dt}|\mathbf{n}_0, t_0)$ as the sum of the probabilities of several mutually exclusive and collectively exhaustive *routes* from $\vec{\mathcal{X}}(t_0) = \mathbf{n}_0$ to $\vec{\mathcal{X}}(t\ +\ \mathrm{dt}) = \mathbf{n}$. These routes are distinguished from one another by the event that happened in the last time interval $[t, t + \mathrm{dt}[$:

$$\begin{aligned} P(\mathbf{n}, t + \mathrm{dt}|\mathbf{n}_0, t_0)\ =\ P(\mathbf{n}, t|\mathbf{n}_0, t_0)\ \times\ \Bigg(1 - \sum_{\boldsymbol{\mu}=1}^{K} \gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\mathbf{n})\,\mathrm{dt}\ +\ o(\,\mathrm{dt})\Bigg)\ + \\ +\ \sum_{\boldsymbol{\mu}=1}^{K} P(\mathbf{n} - \mathbf{s}_{\boldsymbol{\mu}}, t|\mathbf{n}_0, t_0)\ \times\ \Big(\gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\mathbf{n} - \mathbf{s}_{\boldsymbol{\mu}})\,\mathrm{dt}\ +\ o(\,\mathrm{dt})\Big) + \\ +\ o(\,\mathrm{dt})\ . \end{aligned} \tag{4.204}$$

The different routes from $\vec{\mathcal{X}}(t_0) = \mathbf{n}_0$ to $\vec{\mathcal{X}}(t + \mathrm{dt}) = \mathbf{n}$ are obvious from the balance equation (4.204): All routes (i) and (ii) are mutually exclusive since different events are taking place within the last interval $[t, t + \mathrm{dt}[$. The routes subsumed under (iii) can be neglected because they occur with probability of measure zero.

From (4.204) follows straightforwardly the multivariate *chemical master equation*, which is the reference for trajectory simulation: $P(\mathbf{n}, t|\mathbf{n}_0, t_0)$ is subtracted from both sides in equation (4.204), then both sides are divided by dt, the limit $\mathrm{dt} \downarrow 0$ is taken, all $o(\,\mathrm{dt})$ terms vanish and finally we obtain

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\, P(\mathbf{n}, t|\mathbf{n}_0, t_0)\ =\ \sum_{\boldsymbol{\mu}=1}^{K} \Big(\gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\mathbf{n} - \mathbf{s}_{\boldsymbol{\mu}})\, P(\mathbf{n} - \mathbf{s}_{\boldsymbol{\mu}}|\mathbf{n}_0, t_0) - \\ - \gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\mathbf{n})\, P(\mathbf{n}, t|\mathbf{n}_0, t_0)\Big)\ . \end{aligned} \tag{4.205}$$

Initial conditions are required to calculate the time evolution of the probability $P(\mathbf{n}, t|\mathbf{n}_0, t_0)$ and we can easily express them in the form

$$P(\mathbf{n}, t_0|\mathbf{n}_0, t_0) \;=\; \begin{cases} 1\,, & \text{if } \mathbf{n} = \mathbf{n}_0\,, \\ 0\,, & \text{if } \mathbf{n} \neq \mathbf{n}_0\,, \end{cases} \tag{4.205'}$$

which is the same as the sharp initial probability distribution used in the derivation of equation (4.204): $P\big(n_k, t_0|n_k^{(0)}, t_0\big) = \delta_{n_k, n_k^{(0)}}$ for the molecular particle numbers at $t = t_0$. The assumption of extended initial distributions is, of course, also possible but the corresponding master equations become more sophisticated.

4.6.2 Tau-Leaping and higher-level approaches

One general problem of all stochastic simulations with medium and large particle numbers is the enormous consumption of computer time. The necessary but prohibitive amounts of computer capacities are even required when only a single species is present at high particle numbers and this is almost always the case even in the fairly small biological systems within cells. The clear advantage of the stochastic simulation algorithm is at the same time the ultimate cause for its failure to handle most systems in practice: Considering explicitly every single event makes the simulation *exact* but guides it directly into the *computer time requirement trap*.

Tau-leaping. In the section on chemical Langevin equations 4.2.3 τ-leaping has been discussed for justifying the usage of stochastic differential equations in chemical kinetics. Here we mention *τ-leaping* it as an attempt to accelerate the simulation algorithm, which is based on the same idea of lumping together all events happening with a predefined time interval $[t, t+\tau[$ [172, 173]. In contrast to the three implementation of the Monte Carlo step in the original Gillespie simulation algorithm – direct, first-reaction and next-reaction method, which are exact since the consider every event precisely at its time of occurrence – τ-leaping is an approximation whose degree of accuracy depends on the choice of the time interval τ. Assume, for example, τ is chosen so small that only no reaction step or a single reaction step are taking place within the interval $[t, t+\tau[$, then a calculated trajectory obtained by the exact method is indistinguishable for the results of the τ-leaping simulation, which is then exact as well. Choosing a large value of τ, however, will introduce some error that will increase with the size of τ.

The approach is cast into a solid mathematical form through defining a function $\mathcal{P}(\kappa_1, \ldots, \kappa_K|\tau; \mathbf{n}, t[$ which given $\vec{\mathcal{X}}(t) = \mathbf{n}$, measures the probability that exactly κ_j reaction events will occur in the reaction channel $\mathbf{R}_j$ with $j = 1, \ldots, K$. This function $\mathcal{P}$ is the joint probability density of the integer random variables $\mathcal{K}_j(\tau; \mathbf{n}, t)$, which represent the numbers counting how often the reaction channel $\mathbf{R}_j$ is firing in the time interval $[t, t+\tau[$. In order to be able to calculate $\mathcal{P}(\kappa_1, \ldots, \kappa_K|\tau; \mathbf{n}, t)$ with a reasonable effort, an approximation has to be made that determines an appropriate leap size.

Leap condition: The time interval τ has to be chosen so small that none of the K propensity functions $\alpha_j(\mathbf{n}, t); j = 1, \ldots, K$ will change appreciably[67] in the interval $[t, t+\tau[$.

Provided the leap condition is fulfilled and the reaction probability functions remains essentially constant, $\alpha_j(\mathbf{n}) \approx \text{const} \,\forall\, j = 1, \ldots, K$ within the entire time interval, then $\alpha_j(\mathbf{n})\,\mathrm{dt}$ is the probability that a reaction $\mathbf{R}_j$ will take place during any infinitesimal interval dt inside $[t, t+\tau[$, irrespectively

[67] *Appreciably* expresses here the relative change and excludes alterations of macroscopically noninfinitesimal size (see section 4.2.3).

Fig. 4.39 Discrete and continuous time in chemical reactions. .

what happens in the other reaction channels. The events in the individual reaction channels are independent, the random variables follow a Poissonian distribution

$$\mathcal{K}_j(\tau;\mathbf{n},t) \;=\; \pi_{\kappa_j}(\alpha_j\,t) \;=\; (\alpha_j\,t)^{\kappa_j}\,\frac{e^{\alpha_j t}}{\kappa_j!}\;,$$

and this yields for the probability distribution

$$\mathcal{P}(\kappa_1,\ldots,\kappa_K|\tau;\mathbf{n},t) \;=\; \prod_{j=1}^{K}\pi_{\kappa_j}(\alpha_j\,t)\;. \tag{4.206}$$

Each event in the channel $\mathbf{R}_j$ changes the population by $\mathbf{s}_j = \boldsymbol{\nu}'_j - \boldsymbol{\nu}_j$ and accordingly we can easily express the change in the population during the entire interval $[t_i, t_i + \tau_i[$ and the whole trajectory from t_0 to t_N by

$$\boldsymbol{\lambda}_i \;=\; \sum_{j=1}^{K}\kappa_j\,\mathbf{s}_j \quad\text{and}\quad \vec{X}(t_N) \;=\; \vec{\mathcal{X}}(t_0) \;+\; \sum_{i=0}^{N-1}\boldsymbol{\lambda}_i\;, \tag{4.207}$$

where we have already assumed that the leap size τ is variable and can be adjusted to the progress of the reaction.

Tau-leap algorithm: A τ-leap algorithm starts from an initial set of variables, $\vec{\mathcal{X}}(t_0) = \mathbf{n}(t_0) = \mathbf{n}(0)$, then for each $j = 1,\ldots,K$ a sample value κ_j of the random variable $\mathcal{K}_j$ is drawn form the Poissonian $\pi_{\kappa_j}\big(\alpha_j(\mathbf{n}(0),t_0)\big)$ and time and the population vector are advanced by increments, $t_1 = t_0 + \tau$ and $\vec{\mathcal{X}}(t_1) = \mathbf{n}_1 = \mathbf{n}(0) + \boldsymbol{\lambda}_0$. Then progressive iterations, $t_i = t_{i-1} + \tau$ aa well as $\mathbf{n}_i = \mathbf{n}_{i-1} + \boldsymbol{\lambda}_{i-1}$, are performed until one reaches the final time t_N. What is still missing to complete the τ-leap algorithm is an method to determine the loop sizes $\tau_i; i = 0,\ldots,N$. The obvious condition is *effective infinitesimality* of the increments, $|\alpha_j(\mathbf{n}+\boldsymbol{\lambda})-\alpha_j(\mathbf{n})|$, for all reaction channels: $j = 1,\ldots,K$. To find optimal procedures is the major part of the art of doing τ-leap simulations and we refer here to the enormous literature [8, 10, 60, 61, 62, 63, 221, 283, 379, 398, 470] where also many other papers dealing with the choice of the best time interval are found.

The τ-leaping method is not only a valuable computational approach, it can also be seen as providing a link between the chemical master equation (CME) and the chemical Langevin equation (CLE) in the sense that coarse graining of time intervals of size τ is introduced (section 4.2.3).

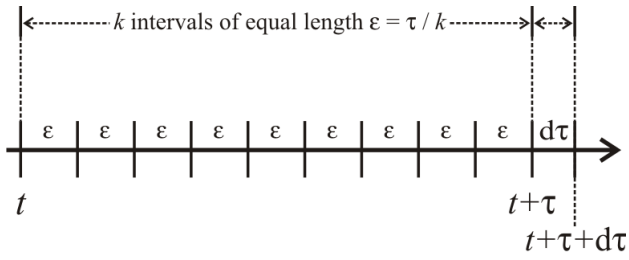


Fig. 4.40 Partitioning of the time interval $[t, t+\tau+\mathrm{d}\tau[$**.** The entire interval is subdivided into $(k+1)$ nonoverlapping subintervals. The first k intervals are of equal size $\varepsilon = \tau/k$ and the $(k+1)$-th interval is of length $\mathrm{d}\tau$.

Hybrid methods. Another class of techniques applied to speeding-up stochastic simulations are hybrid methods [203]. A hybrid algorithm, in essence, treats fast varying variables macroscopically and restricts the stochastic description to the slowly changing particle numbers. Thereby the major computer time wasting part of the algorithm is eliminated: Fast variation of numerically large variables requires an enormously large number of individual jumps. Since fluctuations are relatively small by the $\sqrt{N}$ relation their neglect causes a relatively small error.

4.6.3 The simulation algorithm

The chemical master equation in the form (4.205) is the basis of the simulation algorithm [173] and it is important to realize how the simulation tool fits into the general theoretical framework of the chemical master equation. The simulation algorithm is not based on the probability function $P(\mathbf{n}, t|\mathbf{n}_0, t_0)$ but on another related probability density $p\,(\tau, \boldsymbol{\mu}|\mathbf{n}, t)$, which expresses the probability that given $\vec{\mathcal{X}}(t) = \mathbf{n}$ the *next* reaction in the system will occur in the infinitesimal time interval $[t+\tau, t+\tau+\mathrm{d}\tau[$, *and* it will be an $\mathbf{R}_{\boldsymbol{\mu}}$ reaction.

The probability function $p\,(\tau, \boldsymbol{\mu}|\mathbf{n}, t)$ is the joint density function of two random variables: (i) the *time to the next reaction*, τ, and (ii) the *index of the next reaction*, $\boldsymbol{\mu}$. The possible values of the two random variables are given by the domain of the real variable $0 \leq \tau < \infty$ and the integer variable $1 \leq \boldsymbol{\mu} \leq K$. In order to derive an explicit formula for the probability density $p\,(\tau, \boldsymbol{\mu}|\mathbf{n}, t)$ we introduce the quantity

$$\alpha(\mathbf{n}) \;=\; \sum_{\boldsymbol{\mu}=1}^{K} \gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\mathbf{n})$$

and consider the time interval $[t, t+\tau+\mathrm{d}\tau[$ to be partitioned into $k+1$ subintervals, $k>1$. The first k of these intervals are chosen to be of equal length $\varepsilon = \tau/k$, and together they cover the interval $[t, t+\tau[$ leaving the interval $[t+\tau, t+\tau+\mathrm{d}\tau[$ as the remaining $(k+1)$-th part (figure 4.40). With $\vec{\mathcal{X}}(t) = \mathbf{n}$ the probability $p\,(\tau, \boldsymbol{\mu}|\mathbf{n}, t)$ describes the event *no reaction* occurring in each of the k ε-size subintervals and *exactly one* $\mathbf{R}_{\boldsymbol{\mu}}$ *reaction* in the final infinitesimal $\mathrm{d}\tau$ interval. Making use of conditions 1 and 2 and the multiplication law of probabilities we find

$$p\,(\tau, \boldsymbol{\mu}|\mathbf{n}, t) \;=\; \Big(1 \,-\, \alpha(\mathbf{n})\,\varepsilon \,+\, o(\varepsilon)\Big)^k \Big(\gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\mathbf{n})\,\mathrm{d}\tau \,+\, o(\mathrm{d}\tau)\Big)$$

Dividing both sides by $\mathrm{d}\tau$ and taking the limit $\mathrm{d}\tau \downarrow 0$ yields

$$p\,(\tau, \boldsymbol{\mu}|\mathbf{n}, t) \;=\; \Big(1 \,-\, \alpha(\mathbf{n})\,\varepsilon \,+\, o(\varepsilon)\Big)^k \gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\mathbf{n})$$

This equation is valid for any integer $k>1$ and hence its validity is also guaranteed for $k \to \infty$. Next we rewrite the first factor on the right-hand side of the equation

$$\begin{aligned}\Big(1 \,-\, \alpha(\mathbf{n})\,\varepsilon \,+\, o(\varepsilon)\Big)^k \;&=\; \left(1 \,-\, \frac{\alpha(\mathbf{n})\,k\varepsilon \,+\, k\,o(\varepsilon)}{k}\right)^k \;= \\ &=\; \left(1 \,-\, \frac{\alpha(\mathbf{n})\,\tau \,+\, \tau\,o(\varepsilon)/\varepsilon}{k}\right)^k ,\end{aligned}$$

and take now the limit $k \to \infty$ whereby we make use of the simultaneously occurring convergence $o(\varepsilon)/\varepsilon \downarrow 0$:

$$\lim_{k\to\infty}\Big(1 \,-\, \alpha(\mathbf{n})\,\varepsilon \,+\, o(\varepsilon)\Big)^k \;=\; \lim_{k\to\infty}\left(1 - \frac{\alpha(\mathbf{n})\,\tau}{k}\right)^k \;=\; e^{-\alpha(\mathbf{n})\,\tau}\;.$$

By substituting this result into the initial equation for the probability density of the occurrence of a reaction we find

$$\begin{aligned}p\,(\tau, \boldsymbol{\mu}|\mathbf{n}, t) \;&=\; \alpha(\mathbf{n})\,\frac{\gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\mathbf{n})}{\alpha(\mathbf{n})}\, e^{-\alpha(\mathbf{n})\,\tau} \;= \\ &=\; \gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\mathbf{n})\; e^{-\sum_{\nu=1}^{K}\gamma_{\boldsymbol{\nu}} h_{\boldsymbol{\nu}}(\mathbf{n})\,\tau}.\end{aligned} \tag{4.208}$$

Equation (4.208) provides the mathematical basis for the stochastic simulation algorithm. Given $\vec{\mathcal{X}}(t) = \mathbf{n}$, the probability density consists of two independent probabilities where the first factor describes the *time to the next*

reaction and the second factor the *index of the next reaction*. These factors correspond to two statistically independent random variables η_1 and η_2.

Pseudorandom numbers. In order to implement equation (4.208) for computer simulation we consider probability densities of two unit-interval uniform random variables η_1 and η_2 in order to find the conditions to be imposed on a statistically exact sample pair $(\tau, \boldsymbol{\mu})$: η_1 has an exponential density function with the decay constant $\alpha(\mathbf{n})$,

$$\tau \;=\; \frac{1}{\alpha(\mathbf{n})}\;\ln\bigl(1/\eta_1\bigr)\;, \tag{4.209a}$$

and taking m to be the *smallest* integer which fulfils

$$\boldsymbol{\mu} \;=\; \inf\left\{ m \;\Big|\; \sum_{\mu=1}^{m} \gamma_\mu\, h_\mu(\mathbf{n}) \;>\; \alpha(\mathbf{n})\;\eta_2 \right\}\;. \tag{4.209b}$$

After the values for τ and $\boldsymbol{\mu}$ have been determined the action *advance the state vector* $\vec{\mathcal{X}}(t)$ *of the system* is taking place:

$$\vec{\mathcal{X}}(t) \;=\; \mathbf{n} \quad\longrightarrow\quad \vec{\mathcal{X}}(t+\tau) = \mathbf{n} + \mathbf{s}_{\boldsymbol{\mu}}\;.$$

Repeated application of the advancement procedure is the essence of the stochastic simulation algorithm. It is important to realize that this advancement procedure is exact as far as η_1 and η_2 are obtained by *fair samplings* from a unit interval uniform random number generator or, in other words, the correctness of the procedure depends on the quality of the random number generator applied. Two further issues are important: (i) The algorithm operates with internal time control that corresponds to real time of the chemical process, and (ii) contrary to the situation in differential equation solvers the discrete time steps are not finite interval approximations of an infinitesimal time step and instead, the population vector $\vec{\mathcal{X}}(t)$ maintains the value $\vec{\mathcal{X}}(t) = \mathbf{n}$ throughout the entire finite time interval $[t, t+\mathrm{d}\tau[$ and then changes abruptly to $\vec{\mathcal{X}}(t+\tau) = \mathbf{n} + \mathbf{s}_{\boldsymbol{\mu}}$ at the instant $t+\tau$ when the $\mathbf{R}_{\boldsymbol{\mu}}$ reaction occurs. In other words, there is no *blind interval* during which the algorithm is unable to record changes.

Nonuniformly distributed random numbers. In equation (4.209a) the desired distribution of the pseudorandom variable was built into the expression and the input η_1 was drawn from the uniform distribution. The general approach to derive a continuous random variable with the (cumulative) distribution function $F_{\mathcal{X}}$ is called *inverse transform sampling*: If $\mathcal{X}$ has the distribution $F_{\mathcal{X}}$ then the random variable $\mathcal{Y} = F_{\mathcal{X}}(\mathcal{X})$ is uniformly distributed on the unit interval, $\mathcal{Y}:\; F_{\mathcal{Y}} = \mathcal{U} \in [0,1]$, and this statement can be inverted such that $F_{\mathcal{X}}^{-1}(\mathcal{Y}) = F_{\mathcal{X}}$. The following three step procedure can be used to calculate pseudorandom variables for an invertible distribution function F:

(i) generate a pseudorandom variable u from $\mathcal{U} \in [0,1]$,

Table 4.2 The combinatorial functions $h_{\mu}(\mathbf{n})$ for elementary reactions. Reactions are ordered with respect to reaction order, which in case of mass action is identical to the molecularity of the reaction. Order zero implies that no reactant molecule is involved and the products come from an external source, for example from the influx in a flow reactor. Orders 0, 1, 2, and 3 mean that zero, one, two or three molecules are involved in the elementary step, respectively.

No.	Reaction	Order	$h_{\mu}(\mathbf{n})$
1	$* \longrightarrow$ products	0	1
2	$\mathbf{A} \longrightarrow$ products	1	$n_{\mathbf{A}}$
3	$\mathbf{A} + \mathbf{B} \longrightarrow$ products	2	$n_{\mathbf{A}} n_{\mathbf{B}}$
4	$2\,\mathbf{A} \longrightarrow$ products	2	$n_{\mathbf{A}}(n_{\mathbf{A}} - 1)$
5	$\mathbf{A} + \mathbf{B} + \mathbf{C} \longrightarrow$ products	3	$n_{\mathbf{A}} n_{\mathbf{B}} n_{\mathbf{C}}$
6	$2\,\mathbf{A} + \mathbf{B} \longrightarrow$ products	3	$n_{\mathbf{A}}(n_{\mathbf{A}} - 1) n_{\mathbf{B}}$
7	$3\,\mathbf{A} \longrightarrow$ products	3	$n_{\mathbf{A}}(n_{\mathbf{A}} - 1)(n_{\mathbf{A}} - 2)$

(ii) compute the value $x = F^{-1}(u)$ such that $u = F(x)$, and
(iii) take x as the pseudorandom variable drawn form a distribution given by F.

The procedure is used for the generation of the often required normally distributed pseudorandom numbers and is called Box-Muller transform [50] after the two mathematicians George Edward Pelham Box and Mervin Edgar Muller. Generalizations to discrete variables and arbitrary invertible distribution functions are found in the monograph [89].

Structure of the simulation algorithm. The time evolution of the population in described by the vector $\vec{\mathcal{X}}(t) = \mathbf{n}(t)$, which is updated after every individual reaction event. Reactions are chosen from the set $\mathcal{R} = \{\mathbf{R}_{\boldsymbol{\mu}}; \boldsymbol{\mu} = 1, \ldots, K\}$ defined by the reaction mechanism and the reaction probabilities are contained in a vector $\boldsymbol{\alpha}(\mathbf{n}) = \big(\gamma_1 h_1(\mathbf{n}), \ldots, \gamma_K h_K(\mathbf{n})\big)^{\mathrm{t}}$, which is also updated after every individual reaction event (A classification of reactions according to molecularity is repeated in table 4.2). Updating is performed by adding the stoichiometric vector $\mathbf{s}_{\boldsymbol{\nu}}$ of the chosen reaction $\mathbf{R}_{\boldsymbol{\nu}}$: $\mathbf{n}(t + \mathrm{d}\tau) = \mathbf{n}(t) + \mathbf{s}_{\boldsymbol{\nu}}$ where $\mathbf{s}_{\boldsymbol{\nu}}$ represents a column of the stoichiometric matrix S.

The algorithm comprises five steps:

(i) *Step 0. Initialization*: The time variable is set to $t = 0$, the initial values of all N variables $\mathcal{X}_1, \ldots, \mathcal{X}_N$ for the species – $\mathcal{X}_k$ for species $\mathbf{X}_k$ – are stored, the values for the K parameters of the reactions $\mathbf{R}_{\boldsymbol{\mu}}$, $\gamma_1, \ldots, \gamma_K$, are stored, and the combinatorial expressions are incorporated as factors for the calculation of the reaction rate vector $\boldsymbol{\alpha}(\mathbf{n})$ according to table 4.2 and the probability density $P(\tau, \boldsymbol{\mu})$. Sampling times, $t_1 < t_2 < \cdots$ and the stopping time t_{stop} are specified, the first sampling time is set to

t_1 and stored and the *pseudorandom* number generator is initialized by means of *seeds* or *at random*.

(ii) *Step 1. Monte Carlo step*: A pair of random numbers is created $(\tau, \boldsymbol{\mu})$ by the random number generator according to the joint probability function $P(\tau, \boldsymbol{\mu})$. In essence two explicit methods can be used: the *direct* method and the *first-reaction* method.

(iii) *Step 2. Propagation step*: $(\tau, \boldsymbol{\mu})$ is used to advance the simulation time t and to update the population vector $\mathbf{n}$, $t \to t + \tau$ and $\mathbf{n} \to \mathbf{n} + \mathbf{s}_{\boldsymbol{\mu}}$, then all changes are incorporated in a recalculation of the reaction rate vector $\mathbf{a}$.

(iv) *Step 3. Time control*: Check whether or not the simulation time has been advanced through the next sampling time t_i, and for $t > t_i$ send current t and current $\mathbf{n}(t)$ to the output storage and advance the sampling time, $t_i \to t_{i+1}$. Then, if $t > t_{\text{stop}}$ or if no more reactant molecules remain leading to $h_{\boldsymbol{\mu}} = 0 \; \forall \; \boldsymbol{\mu} = 1, \ldots, K$, finalize the calculation by switching to *step 4*, and otherwise continue with *step 1*.

(v) *Step 4. Termination*: Prepare for final output by setting flags for early termination or other unforseen stops and send final time t and final $\mathbf{n}$ to the output storage and terminate the computation.

A caveat is needed for the integration of stiff systems where the values of individual variable can vary by many orders of magnitude and such a situation might caught the calculation in a trap by slowing down time progress.

The Monte Carlo step. Pseudorandom numbers are drawn from a random number generator of sufficient quality whereby quality is meant in terms of no or very long recurrence cycles and a the closeness of the distribution of the pseudorandom numbers r to the uniform distribution on the unit interval:

$$0 \leq a < b \leq 1 \qquad \Longrightarrow \qquad P(a \leq \eta \leq b) \; = \; b - a \; .$$

With this prerequisite we discuss now two methods which use two output values η of the pseudorandom number generator to generate a random pair $(\tau, \boldsymbol{\mu})$ with the prescribed probability density function $P(\tau, \boldsymbol{\mu})$.

The ***direct*** *method.* The two-variable probability density is written as the product of two one-variable density functions:

$$P(\tau, \boldsymbol{\mu}) \; = \; P_1(\tau) \cdot P_2(\boldsymbol{\mu}|\tau) \; .$$

Here, $P_1(\tau)\,\mathrm{d}\tau$ is the probability that the next reaction will occur between times $t + \tau$ and $t + \tau + \mathrm{d}\tau$, irrespective of which reaction it might be, and $P_2(\boldsymbol{\mu}|\tau)$ is the probability that the next reaction will be an $\mathbf{R}_{\boldsymbol{\mu}}$ given that the next reaction occurs at time $t + \tau$.

By the addition theorem of probabilities, $P_1(\tau)\,\mathrm{d}\tau$ is obtained by summation of $P(\tau, \boldsymbol{\mu})\,\mathrm{d}\tau$ over all reactions $\mathbf{R}_{\boldsymbol{\mu}}$:

$$P_1(\tau) \;=\; \sum_{\boldsymbol{\mu}=1}^{K} P(\tau, \boldsymbol{\mu}) \;. \tag{4.210}$$

Combining the last two equations we obtain for $P_2(\boldsymbol{\mu}|\tau)$

$$P_2(\boldsymbol{\mu}|\tau) \;=\; P(\tau, \boldsymbol{\mu}) \;/\; \sum_{\boldsymbol{\nu}}^{K} P(\tau, \boldsymbol{\nu}) \tag{4.211}$$

Equations (4.210) and (4.211) express the two one-variable density functions in terms of the original two-variable density function $P(\tau, \boldsymbol{\mu})$. From equation (4.208) we substitute into $P(\tau, \boldsymbol{\mu}) = p\,(\tau, \boldsymbol{\mu}|\mathbf{n}, t)$ through simplifying the notation by using

$$\alpha_{\boldsymbol{\mu}} \;\equiv\; \gamma_{\boldsymbol{\mu}} h_{\boldsymbol{\mu}}(\mathbf{n}) \;\text{ and }\; \alpha \;=\; \sum_{\boldsymbol{\mu}=1}^{K} \alpha_{\boldsymbol{\mu}} \;\equiv\; \sum_{\boldsymbol{\mu}=1}^{K} \gamma_{\boldsymbol{\mu}} h_{\boldsymbol{\mu}}(\mathbf{n})$$

and find

$$\begin{aligned} P_1(\tau) \;&=\; \alpha \exp\left(-\alpha\,\tau\right), \quad 0 \le \tau < \infty \;\text{ and} \\ P_2(\boldsymbol{\mu}|\tau) \;&=\; P_2(\boldsymbol{\mu}) \;=\; \alpha_{\boldsymbol{\mu}} \;/\; \alpha\,, \quad \boldsymbol{\mu} = 1, \ldots, K \;. \end{aligned} \tag{4.212}$$

As indicated, in this particular case, $P_2(\boldsymbol{\mu}|\tau)$ turns out to be independent of τ. Both one variable density functions are properly normalized over their domains of definition:

$$\int_0^{\infty} P_1(\tau)\,\mathrm{d}\tau \;=\; \int_0^{\infty} \alpha\, e^{-\alpha\,\tau}\,\mathrm{d}\tau \;=\; 1 \;\text{ and }\; \sum_{\boldsymbol{\mu}=1}^{K} P_2(\boldsymbol{\mu}) \;=\; \sum_{\boldsymbol{\mu}=1}^{K} \frac{\alpha_{\boldsymbol{\mu}}}{\alpha} \;=\; 1 \;.$$

Thus, in the *direct* method a random value τ is created from a random number on the unit interval, η_1, and the distribution $P_1(\tau)$ by taking

$$\tau \;=\; -\,\frac{\ln \eta_1}{\alpha} \;. \tag{4.213}$$

The second task is to generate a random integer $\widehat{\boldsymbol{\mu}}$ according to $P_2(\boldsymbol{\mu}|\tau)$ in such a way that the pair $(\tau, \boldsymbol{\mu})$ will be distributed as prescribed by $P(\tau, \boldsymbol{\mu})$. For this goal another random number, η_2, will be drawn from the unit interval and then $\widehat{\boldsymbol{\mu}}$ is taken to be the integer that fulfils

$$\sum_{\boldsymbol{\nu}=1}^{\boldsymbol{\mu}-1} \alpha_{\boldsymbol{\nu}} \;<\; \eta_2\,\alpha \;\le\; \sum_{\boldsymbol{\nu}=1}^{\boldsymbol{\mu}} \alpha_{\boldsymbol{\nu}} \;. \tag{4.214}$$

The values α_1, α_2, ... , are cumulatively added in sequence until their sum is observed to be equal or to exceed $\eta_2\alpha$ and then $\hat{\boldsymbol{\mu}}$ is set equal to the

index of the last $\alpha_{\boldsymbol{\nu}}$ term that had been added. Rigorous justifications for equations (4.213) and (4.214) are found in [166, pp.431-433]. If a fast and reliable uniform random number generator is available, the *direct* method can be easily programmed and rapidly executed. This it represents a simple, fast, and rigorous procedure for the implementation of the *Monte Carlo* step of the simulation algorithm.

The ***first-reaction*** *method.* This alternate method for the implementation of the *Monte Carlo* step of the simulation algorithm is not quite as efficient as the *direct* method but it is worth presenting here because it adds insight into the stochastic simulation approach. Adopting again the notation $\alpha_{\boldsymbol{\nu}} \equiv \gamma_{\boldsymbol{\nu}} h_{\boldsymbol{\nu}}(\mathbf{n})$ it is straightforward to derive

$$P_{\boldsymbol{\nu}}(\tau)\,\mathrm{d}\tau \;=\; \alpha_{\boldsymbol{\nu}}\,\exp\left(-\alpha_{\boldsymbol{\nu}}\,\tau\right)\mathrm{d}\tau \tag{4.215}$$

from (4.199) and (4.200). Then, $P_{\boldsymbol{\nu}}(\tau)$ would indeed be the probability at time t for an $\mathbf{R}_{\boldsymbol{\nu}}$ reaction to occur in the time interval $[t+\tau, t+\tau+d\tau[$ were it not for the fact that the number of $\mathbf{R}_{\boldsymbol{\nu}}$ reactant combinations might have been altered between t and $t+\tau$ by the occurrence of other reactions. Taking this into account, a *tentative reaction time* $\tau_{\boldsymbol{\nu}}$ for $\mathbf{R}_{\boldsymbol{\nu}}$ is generated according to the probability density function $P_{\boldsymbol{\nu}}(\tau)$, and in fact, the same can be done for all reactions $\{\mathbf{R}_{\boldsymbol{\mu}}\}$. We draw a random number $\eta_{\boldsymbol{\nu}}$ from the unit interval and compute

$$\tau_{\boldsymbol{\nu}} \;=\; -\frac{\ln \eta_{\boldsymbol{\nu}}}{\alpha_{\boldsymbol{\nu}}}\,,\;\; \boldsymbol{\nu} = 1,\ldots,K\;. \tag{4.216}$$

From these K *tentative next* reactions the one, which occurs first, is chosen to be the *actual next* reactions:

$$\begin{aligned} \tau \;&=\; \text{smallest } \tau_{\boldsymbol{\nu}} \;\text{ for all }\; \boldsymbol{\nu} = 1,\ldots,K\;, \\ \boldsymbol{\mu} \;&=\; \boldsymbol{\nu} \text{ for which } \tau_{\boldsymbol{\nu}} \text{ is smallest}\,. \end{aligned} \tag{4.217}$$

Daniel Gillespie [166, pp.420-421] provides a straightforward proof that the random $(\tau, \boldsymbol{\mu})$ obtained by the *first reaction* method is in full agreement with the probability density $P(\tau, \boldsymbol{\mu})$ from equation (4.208).

It is tempting to try to extend the *first reaction* methods by letting the *second next* reaction be the one for which $\tau_{\boldsymbol{\nu}}$ has the second smallest value. This, however, is in conflict with correct updating of the vector of particle numbers, $\mathbf{n}$, because the results of the first reaction are not incorporated into the combinatorial terms $h_{\boldsymbol{\mu}}(\mathbf{n})$. Using the second earliest reaction would, for example, allow the second reaction to involve molecules already destroyed in the first reaction but would not allow the second reaction to involve molecules created in the first reaction.

The ***next-reaction*** *method.* In the next reaction method one makes use of all reaction times $\tau_{\boldsymbol{\nu}}$ that were calculated for the next reaction step as in the first reaction method [164]. Three expensive actions taking time proportional

to the number of reactions K are performed in every iteration step of the first-reaction method: (i) updating all K values $\alpha_{\boldsymbol{\mu}}$, (ii) generating a putative reaction time $\tau_{\boldsymbol{\nu}}$ for every $\boldsymbol{\nu}$, and (iii) identifying the shortest putative $\tau_{\boldsymbol{\mu}}$. The next-reaction method is somewhat more involved but avoids the time-wasting calculations of future random times that are not used after the reaction event has occurred, it has been proven to be exact in the same sense as he direct and the first-reaction method.

The basic idea of the next-reaction method is to reuse the already calculated times $\tau_{\boldsymbol{\nu}}$ wherever this is appropriate. There is however one important caveat: Monte Carlo simulations, in general, assume random numbers that are statistically independent, and therefore reuse of random numbers and quantities derived from them is illegitimate. In the specific case of the next-reaction algorithm, however, it has been verified by proof that all putative reactions times can be reused except the time $\tau_{\boldsymbol{\mu}}$ of the reaction, which was executed. By means of a *dependency graph* that follows directly from the reaction mechanism only the minimal number of $\alpha_{\boldsymbol{\mu}}$-values are updated. Storage of all $\tau_{\boldsymbol{\nu}}$-times together with the $\alpha_{\boldsymbol{\nu}}$-values is required and efficient implementations of the next-reaction are using special data structures [164].

Thus, the *first-reaction* method is just as rigorous as the *direct* method and it is probably easier to implement in a computer code than the direct method. From a computational efficiency point of view, however, the direct method is preferable because for $K \geq 3$ it requires fewer random numbers and hence the first reaction methods is wasteful. This question of economic use of computer time is not unimportant because stochastic simulations in general are taxing the random number generator quite heavily. For $K \geq 3$ and in particular for large K the direct method is probably the method of choice for the Monte Carlo step. The *next-reaction* method is exact too and can be seen as a more efficient extension of the first reaction method, which for sufficiently large M and K beats also the direct method in efficiency, because it requires asymptotically only one random number per reaction event.

Computer codes. An early computer code of the simple version of the algorithm described – still in *FORTRAN* – is found in [166]. Meanwhile many attempts were made in order to speed-up computations and allow for simulation of stiff systems (see e.g. [60]. A recent review of the simulation methods also contains a discussion of various improvements of the original code [173].

Several computer codes in different languages including C++ are now available on the internet and unless one aims at an efficient program for some special task it does not pay to write another code unless for educational purposes. The simulations reported here were performed with a *Mathematica 7,8* implementation that runs with small modifications also under *Mathematica 9*. Other equally efficient and user friendly implementations are available for *Matlab* and other high-level user interfaces. A didactic introduction is found in [207] and sample programs for Matlab are available from `http:personal.strath.ac.uk/d.j.higham/algfiles.html`.

xCellerator project [407]
StochKit [281] [388]

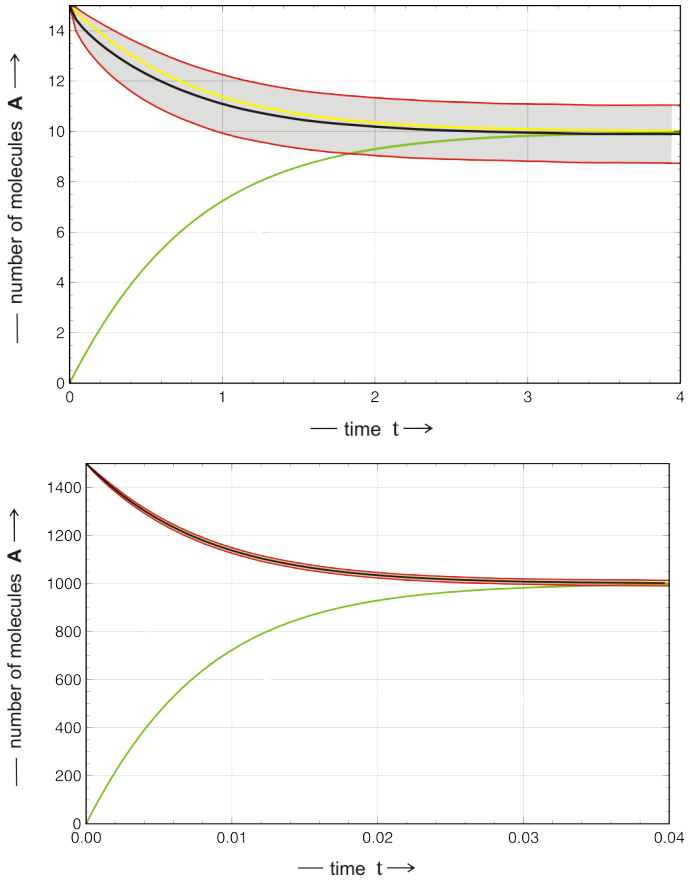


Fig. 4.41 The reversible bimolecular reaction $\mathbf{A}+\mathbf{B} \rightleftharpoons 2\mathbf{C}$. The figure shows the expectation value $\mathrm{E}\big((n_{\mathsf{A}}(t)\big)$ (black) embedded between the curves $\mathrm{E}(n_{\mathsf{A}}) \pm \sigma(n_{\mathsf{A}})$ (red) together with the deterministic solution $\widehat{n}_{\mathsf{A}}(t) = a(t)$ (yellow) and $\widehat{n}_{\mathsf{C}} = c(t)$ (green). Parameter choice: $k =$, $l =$, $n_{\mathsf{C}}(0) = 0$, and $n_{\mathsf{A}}(0) = 15$, $n_{\mathsf{B}}(0) =$ (upper plot) and $n_{\mathsf{A}}(0) = 1500$, $n_{\mathsf{B}}(0) =$ (lower plot).

4.6.4 Examples of simulations

In this section we shall be dealing with some selected examples of numerical simulations using the Gillespie algorithm.

Reversible bimolecular reaction $\mathbf{A}+\mathbf{B} \rightleftharpoons 2\mathbf{C}$. Recalling our results from section 4.3.3 the major obstacle for driving analytical solutions was the tridiagonal structure of the transition matrix W that did not allow for the derivation of analytical expressions for the eigenvalues, and we could only derive results for the equilibrium distribution. In order to be able to compare a process as

close as possible to the simple autocatalytic reaction $\mathbf{A}+\mathbf{X} \rightleftharpoons 2\mathbf{X}$ with choose here $\mathbf{A}+\mathbf{B} \rightleftharpoons 2\mathbf{C}$.

The results of the computations with two different sample sizes are shown in figure 4.41. Starting from sharp initial conditions at $t = t_0$, $P_n(0) = \delta_{n,n_0}$ and with small particle numbers, $n_\mathsf{A} = n = 15$, the one standard deviation band increases in width until it reaches the equilibrium value. The deterministic solution $\widehat{n}(t)$ differs slightly but significantly from the expectation value $\mathrm{E}\big(n(t)\big)$, which is in full agreement with the previous analysis since the first jump moment α_1 is not linear in n. The deviation between the two curves is readily explained: The stochastic reverse reaction, $2\mathbf{C} \to \mathbf{A}+\mathbf{B}$ fulfils the rate function $\chi_{\leftarrow}^{(MA)}(\mathbf{n}) = \gamma h(\mathbf{n}) = l\, n_\mathsf{C}(n_\mathsf{C}-1)$ in mass action kinetics and is slower than the deterministic rate function $v_{\leftarrow}^{(MA)} = l\, n_\mathsf{C}^2$ (see figure 4.23). The stochastic and deterministic rate functions of the reversible reaction are:

$$\begin{aligned}
\chi_{\rightarrow}^{(MA)} - \chi_{\leftarrow}^{(MA)} &= -k\, n_\mathsf{A} \cdot n_\mathsf{B} + l\, n_\mathsf{C}^2 - l\, n_\mathsf{C} \quad \text{and} \\
v_{\rightarrow}^{(MA)} - v_{\leftarrow}^{(MA)} &= -k\, n_\mathsf{A} \cdot n_\mathsf{B} + l\, n_\mathsf{C}^2 \; .
\end{aligned}$$

The deterministic rate is smaller and hence the curve $\widehat{n}(t)$ is flatter. At sufficiently high particle numbers all four curves, $\widehat{n}$, $\mathrm{E}(n)$, $\mathrm{E}(n) \pm \sigma(n)$, almost coincide within the thickness of the lines again in full agreement with the expectation $\sigma \propto \sqrt{n}$.

The extended Michaelis-Menten reaction. The extended mechanism of Michaelis-Menten type enzyme catalysis (figure 4.2, version A)

$$\mathbf{S} + \mathbf{E} \underset{l_1}{\overset{k_1}{\rightleftharpoons}} \mathbf{SE} \underset{l_2}{\overset{k_2}{\rightleftharpoons}} \mathbf{EP} \underset{l_3}{\overset{k_3}{\rightleftharpoons}} \mathbf{E} + \mathbf{P}$$

is dealing with five species involved in three reaction steps, and accordingly we have two conservation relations:

$$s_0 + p_0 = s + c + d + p \quad \text{and} \quad e_0 = e + c + d \; ,$$

with $s = [\mathbf{S}]$, $e \equiv [\mathbf{E}]$, $p = [\mathbf{P}]$, $c \equiv [\mathbf{SE}]$, and $d \equiv [\mathbf{PE}]$.

Autocatalysis and fluctuations. Analytical results for the irreversible autocatalytic reaction $\mathbf{A}+\mathbf{X} \to 2\mathbf{X}$ have been discussed already in section 4.3.3.3. Here, we present and analyze the results of numerical simulations of the reversible autocatalytic process $\mathbf{A}+\mathbf{X} \rightleftharpoons 2\mathbf{X}$ in order to work out the difference between ordinary and autocatalytic processes. One feature is particularly evident and can be explained easily: When starting from the comparable initial conditions, stochastic process requires more time than the conventional ODE solution in the approach towards equilibrium. Although the deterministic curve lies within the one standard deviation-zone, $\mathrm{E}(t) \pm \sigma(t)$, the convergence of the stochastic solution towards the deterministic limit with increasing num-

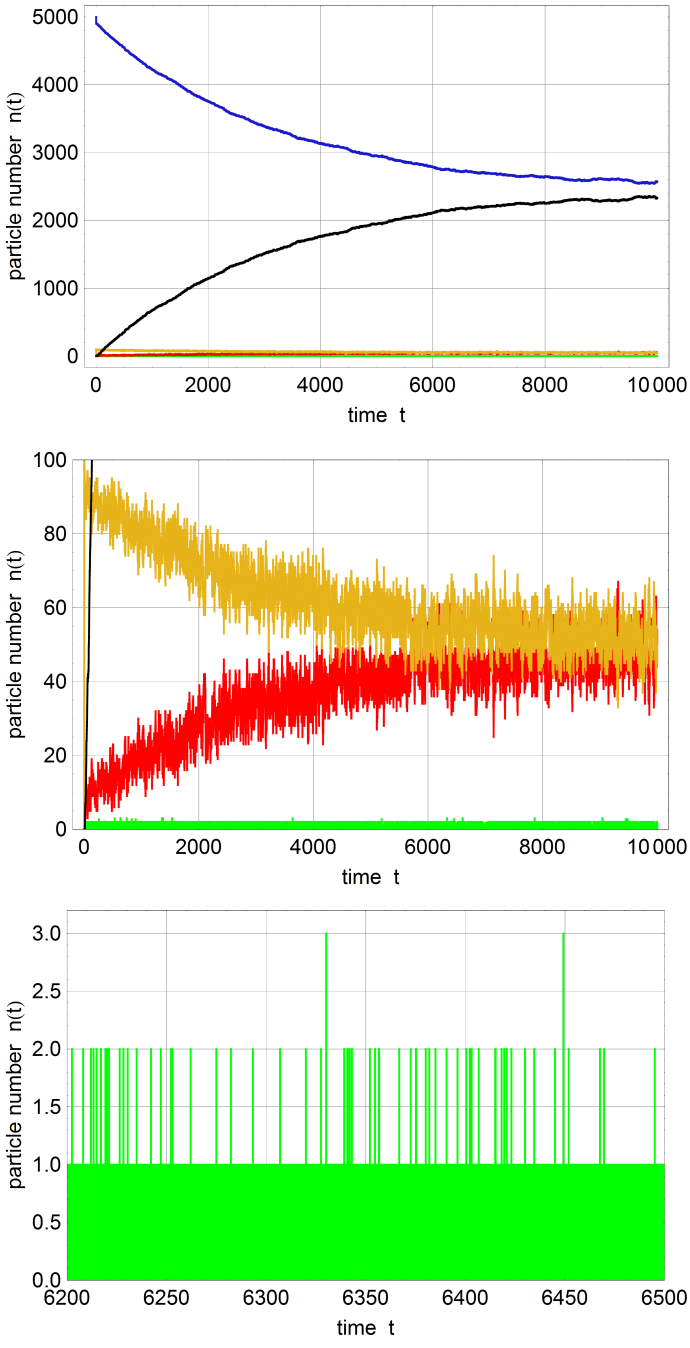


Fig. 4.42 Continued on next page.

Fig. 4.42 The extended Michaelis-Menten reaction. The fully reversible mechanism shown as version A in figure 4.2 is simulated in form of a single trajectory with large excess of substrate. The plot at the top shows the number of substrate and product molecules, $s(t)$ (blue), and product, $p(t)$ (black), the plot in the middle presents particle numbers for the two complexes $[\mathbf{S}\cdot\mathbf{E}] = c(t)$ (yellow) and $[\mathbf{E}\cdot\mathbf{P}] = d(t)$ (red), and the bottom plot eventually shows the number of free enzyme molecules, $e(t)$, which almost always takes on only four different values: $e \in \{0, 1, 2, 3\}$.

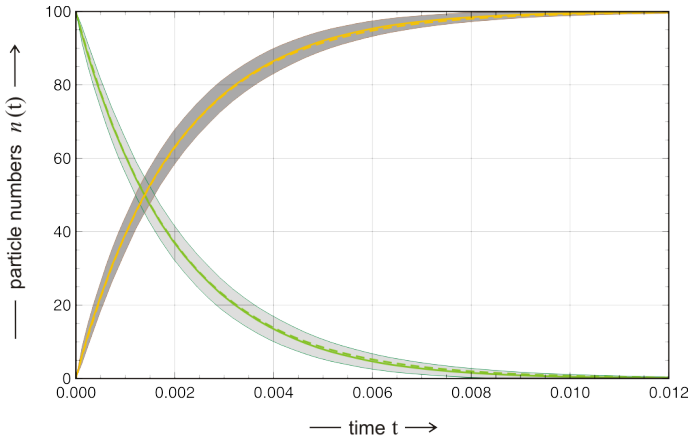
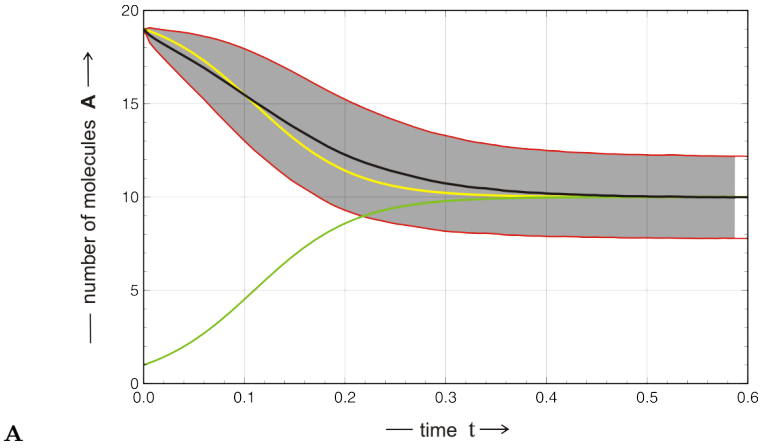


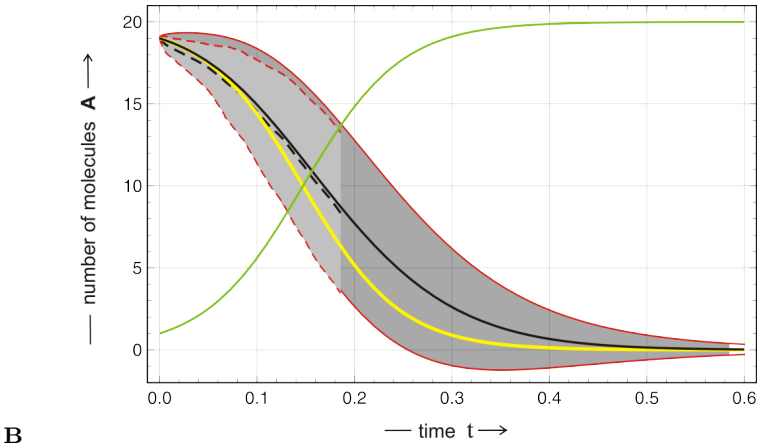
Fig. 4.43 Enzyme-substrate binding. The binding step preceding the enzymatic reaction is assumed to be faster than the conversion of substrate into product.

bers of molecules, n_0, is not evident for population sizes up to $N = 10\,000$. In figure 4.44 we present the autocatalytic process for the comparison of the stoichiometrically closely related bimolecular reaction$\mathbf{A} + \mathbf{B} \rightleftharpoons 2\mathbf{C}$ discussed above. The expectation values of the stochastic processes show the same qualitative behavior as the solutions of conventional kinetics shown in figure 5.1:

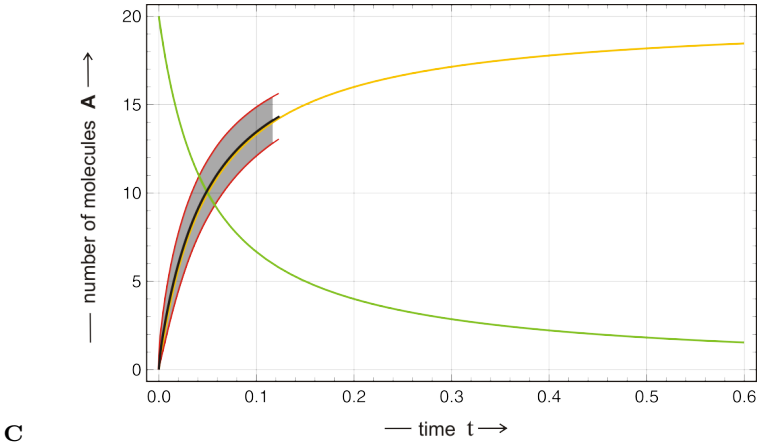
The uncatalyzed reversible process $\mathbf{A} + \mathbf{B} \rightleftharpoons 2\mathbf{C}$ shows the well-known hyperbolic approach towards stationarity whereas self-enhancement of autocatalysis in $\mathbf{A} + \mathbf{X} \rightleftharpoons 2\mathbf{X}$ leads to *sigmoid* or *S-shaped* curves (figure 4.44). The difference between the two processes becomes even more apparent when fluctuation are taken into account: The remarkable effect of fluctuation enhancement in autocatalysis broadens substantially the one-standard deviation envelope: At $\bar{n} \approx 10$ we observe a band in the autocatalytic process that is approximately twice as broad as the one in the uncatalyzed reaction. At larger particle numbers $\bar{n} \approx 1000$ the difference is dramatic: The one standard deviation band is still large in the autocatalytic reaction whereas fluctuation have been reduced tenfold and became very small autocatalysis-free system.



A

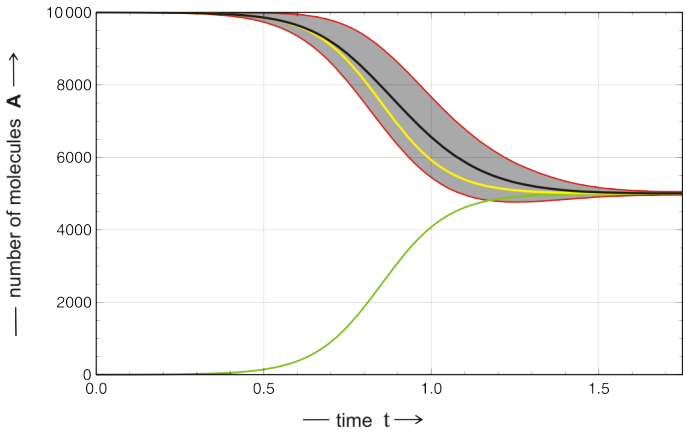


B

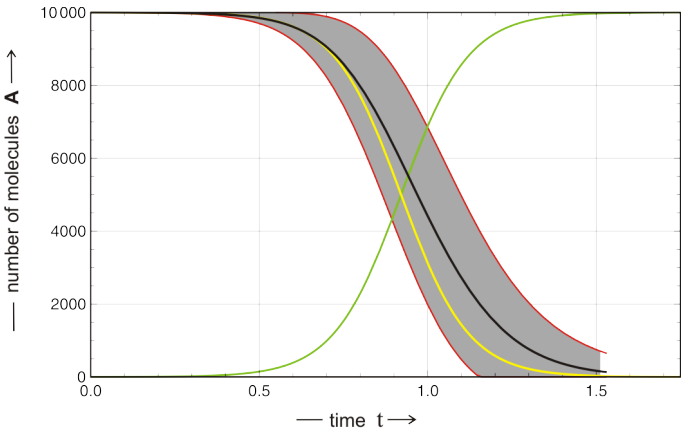


C

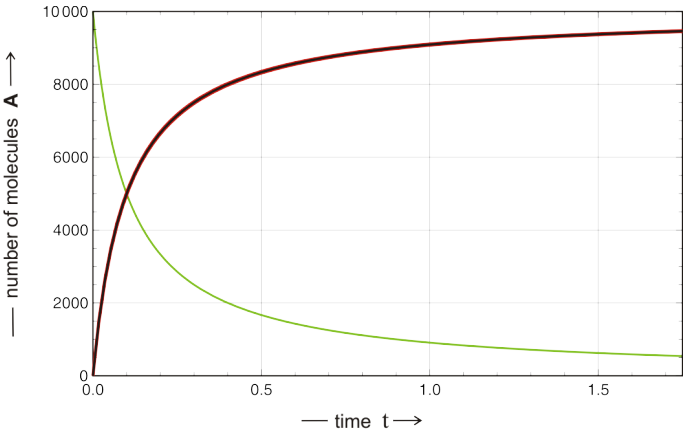
Fig. 4.44 Continued on next page.



D



E



F

Fig. 4.44 Continued on next page.

Fig. 4.44 The role of fluctuations in autocatalytic reactions. The figure consists of six individual plots derived from the autocatalytic reaction $\mathsf{A}+\mathsf{X} \rightleftharpoons 2\mathsf{X}$. The first three plots (**A**,**B**, and **C**) show the reversible reaction and the $\mathbf{R}_{\leftrightarrow}$ together with the two irreversible reaction steps, $\mathbf{R}_{\rightarrow}$ and $\mathbf{R}_{\leftarrow}$, with total particle numbers $N = n_{\mathsf{A}} + n_{\mathsf{X}} = 20$ and the other three plots (**D**,**E**, and **F**) show the tree reactions, $\mathbf{R}_{\leftrightarrow}$, $\mathbf{R}_{\rightarrow}$ and $\mathbf{R}_{\leftarrow}$, in a population of one hundredfold size, $N = 2000$. In plot **B** the analytical solution is compared with computer simulations performed with the Gillespie algorithm (broken lines and light gray area).
Parameter choice: $k = l = 1$ mole^{-1}l^{-1}sec^{-1}, initial conditions: ($n_{\mathsf{A}}(0) = 19$, $n_{\mathsf{X}}(0) = 1$) or $n_{\mathsf{X}}(0) = 20$ and ($n_{\mathsf{A}}(0) = 9999$, $n_{\mathsf{X}}(0) = 1$) or $n_{\mathsf{X}}(0) = 10000$, color code: $\mathrm{E}\big(n_{\mathsf{A}}(t)\big)$ black, $\mathrm{E} \pm \sigma\big(n_{\mathsf{A}}(t)\big)$ red, $\widehat{n}_{\mathsf{A}}(t)$ yellow, and $\widehat{n}_{\mathsf{C}}(t)$ green.

In order to learn more about the origin of the fluctuation enhancement the reversible reaction, $\mathbf{R}_{\leftrightarrow}$, was resolved into the two irreversible steps,

$$\mathbf{R}_{\rightarrow}\colon\ \mathsf{A}+\mathsf{X} \rightarrow 2\mathsf{X} \quad \text{and} \quad \mathbf{R}_{\leftarrow}\colon 2\mathsf{X} \rightarrow \mathsf{A}+\mathsf{X}\ .$$

The forward reaction step $\mathbf{R}_{\rightarrow}$ shows the characteristic features of the autocatalytic process whereas the reverse reaction $\mathbf{R}_{\leftarrow}$ resembles a normal non-autocatalytic reaction where the fluctuation obey the $\sqrt{n}$-law of the fluctuation-dissipation theorem. Not unexpectedly it is the $\mathsf{X} \rightarrow 2\mathsf{X}$ element of the process that gives rise to self-enhancement.

Eventually we mention that the autocatalytic reaction with buffered concentration of $[\mathsf{A}] = a_0$, corresponding to an open system

$$\mathsf{X} \quad \xrightarrow{ka_0} \quad 2\,\mathsf{X}\ , \tag{4.218}$$

has been studied already by Max Delbrück [87]. The stochastic process is identical to a simple birth process with birth rate $\lambda \cdot n = ka_0$ and will be discussed in the context of other birth-and in section 5.2.1. Enhancement of fluctuations to macroscopic level is observed as a characteristic for unconstrained autocatalytic growth.

Higher order autocatalysis: Bistability and Oscillations. So far all chemical reactions were approaching either a unique thermodynamic equilibrium or a stationary state depending on the embedding of the system giving rise to a closed od an open state, respectively. First order autocatalytic systems exhibited some features that are otherwise uncommon in chemical kinetics, the characteristic example is self-enhancement of fluctuations. Here we consider more prominent nonlinear phenomena in chemistry, two or more stable steady states and oscillations [387]. An example of a simple mechanism showing bistability consists of a embedding of the termolecular autocatalytic reaction step (4.1m), $\mathsf{A}+2\mathsf{X} \leftrightharpoons 3\mathsf{X}$ in the flow reactor. In order to be consistent with chemical thermodynamics the uncatalyzed $\mathsf{A} \leftrightharpoons \mathsf{X}$ is added. The occurrence of oscillations in concentrations is demonstrated and analyzed of

two chemical model systems, the Brusselator and the Oregonator. The former mechanism was conceived as a simple model that allows for the occurrence of oscillations and spatial pattern formation in a chemical reaction network whereas the latter was postulated as a simplified model mechanism of the Belousov-Zhabotinsky reaction. We use both systems here for the purpose of illustrating the influence of stochasticity on complex dynamics, in particular on simple bifurcations.

As mentioned in the introduction of chemical elementary step reactions termolecular reaction steps are based on highly improbable encounters of three molecules and therefore excluded in conventional reaction kinetics and indeed, the fully resolved multistep mechanisms of higher order autocatalytic reactions involve only mono and bimolecular steps. We mention in this context a beautiful mathematical exercise consisting of the tasks to find the smallest reaction systems exhibiting oscillation resulting from a Hopf bifurcation[68] [461] or showing bistability [460].

Bistabile reaction networks. The termolecular autocatalytic reaction step (4.1m) together with the corresponding uncatalyzed reaction in the flow reactor give rise to the mechanism:

$$\star \xrightarrow{r\cdot a_0} \mathbf{A}\,, \tag{4.219a}$$

$$\mathbf{A} \underset{\kappa\, l_1}{\overset{\kappa\, k_1}{\rightleftharpoons}} \mathbf{X}\,, \tag{4.219b}$$

$$\mathbf{A} + 2\,\mathbf{X} \underset{l_1}{\overset{k_1}{\rightleftharpoons}} 3\,\mathbf{X}\,, \tag{4.219c}$$

$$\mathbf{A} \xrightarrow{r} \oslash \tag{4.219d}$$

$$\mathbf{X} \xrightarrow{r} \oslash\,, \tag{4.219e}$$

which corresponds to an overall conversion of $\mathbf{A}$ into $\mathbf{X}$. The kinetic differential equations, $[\mathbf{A}] = a$ and $[\mathbf{X}] = x$,

$$\begin{aligned} \frac{\mathrm{d}a}{\mathrm{d}t} &= -(k_1\, a - l_1\, x)(\kappa + x^2) + r(a_0 - a) \quad \text{and} \\ \frac{\mathrm{d}x}{\mathrm{d}t} &= +(k_1\, a - l_1\, x)(\kappa + x^2) - r\, x\,, \end{aligned} \tag{4.220}$$

[68] The Hopf or Poincaré-Andronov-Hopf bifurcation is named after Henri Poincaré, the German-US-American mathematician Eberhard Hopf and the Russian physicist Aleksandr Andronov, and occurs, in essence, when a complex conjugate pair of eigenvalues crosses the real axis, $\lambda_{1,2} = \alpha \pm \beta$ and $\alpha < 0 \implies alpha > 0$ [404, p.48 ff.].

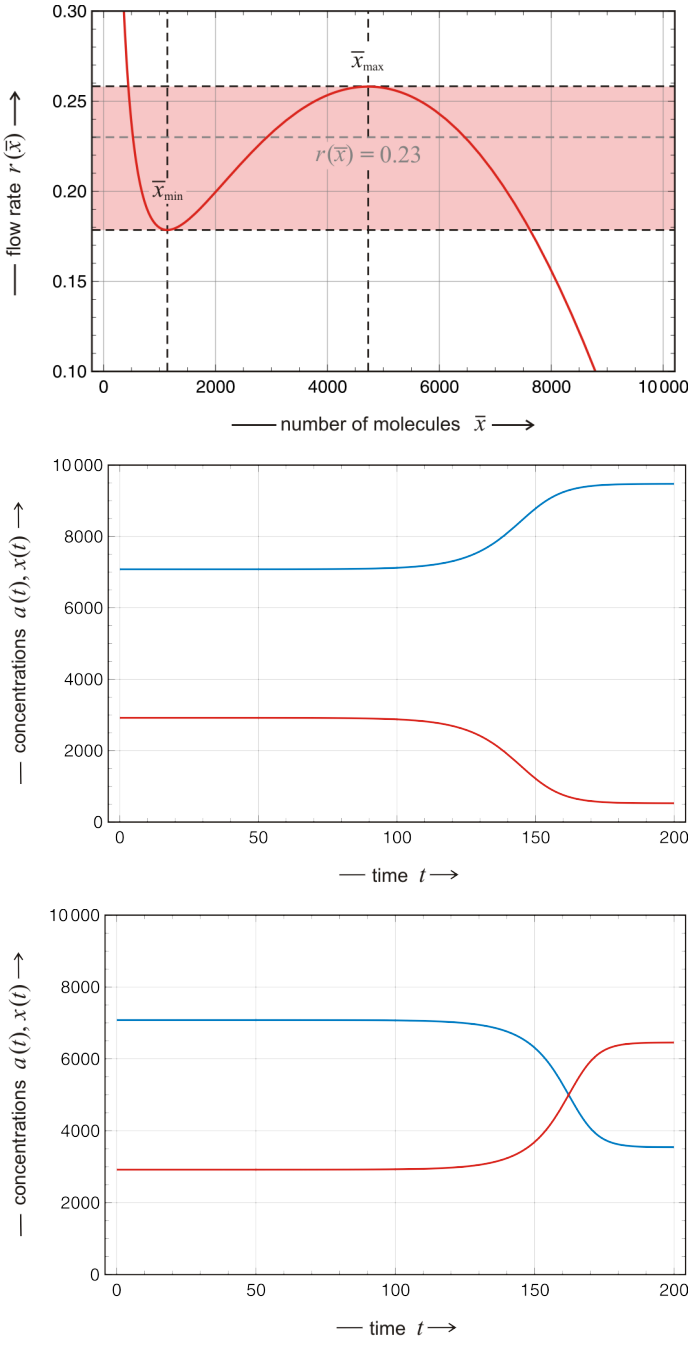


Fig. 4.45 Continued on next page.

Fig. 4.45 Analysis of bistability in chemical reaction networks. The reaction mechanism (4.219) sustains three stationary states, $S_1(\bar{x} = \bar{x}_1)$, $S_2(\bar{x} = \bar{x}_2)$, and $S_3(\bar{x} = \bar{x}_3)$, in the range $r(\bar{x}_{\min}) < r(\bar{x}) < r(\bar{x}_{\max})$ with $\bar{x}_1 < \bar{x}_2 < \bar{x}_3$. The two states S_1 and S_3 are asymptotically stable and S_2 is an unstable saddle point. The plot in the middle shows the solution curves $a(t)$ (blue) and $x(t)$ (red) starting from initial conditions just above the unstable state, $x(0) = \bar{x}_2 - \delta$, and the system converges to state S_1. Analogously, the plot at the bottom starts at $x(0) = \bar{x}_2 + \delta$ and the trajectory ends at state S_1. Parameter choice: $k_1 = 1.0 \times 10^{-10}\,[t^{-1} \cdot \mathrm{M}^{-2}]$, $l_1 = 1.0 \times 10^{-8}\,[t^{-1} \cdot \mathrm{M}^{-2}]$, $\kappa = 10^6\,[\mathrm{M}^2]$, $a_0 = 10\,000\,[\mathrm{M}]$ and $r = 0.23\,[t^{-1}]$. Steady state concentrations: $\bar{x}_1 = 525.34\,[\mathrm{M}]$, $\bar{x}_2 = 2918.978\,[\mathrm{M}]$, and $\bar{x}_3 = 6456.67\,[\mathrm{M}]$. Initial conditions: $x(0) = 2918.97\,[\mathrm{M}]$ (middle plot) and $x(0) = 2918.98\,[\mathrm{M}]$ (bottom plot).

lead to $\mathrm{d}(a+x)/\,\mathrm{dt} = \mathrm{d}a/\,\mathrm{dt} + \mathrm{d}x/\,\mathrm{dt} = r\big(a_0 - (a+x)\big)$ with the stationary solution $\bar{a} + \bar{x} = a_0$ and sustain one or three steady states $(\bar{a} = a_0 - \bar{x}, \bar{x})$ that fulfil the implicit equation [368, pp.18-27]

$$r(\bar{x}) \;=\; \frac{1}{\bar{x}}\Big(\kappa k_1 a_0 \;-\; \bar{x}\,\kappa(k_1 + l_1) \;+\; \bar{x}^2\,k_1 a_0 \;-\; \bar{x}^3\,(k_1 + l_1)\Big)\;. \qquad (4.221)$$

In case we are dealing with three stationary states, $S_1(\bar{x} = \bar{x}_1)$, $S_2(\bar{x} = \bar{x}_2)$, and $S_3(\bar{x} = \bar{x}_3)$, S_1 and S_3 are asymptotically stable and the saddle S_2 separates the two basins of attraction, $x < \bar{x}_2$ and $x > \bar{x}_2$, respectively. The subdomains with one or three real and positive solutions for $\bar{x}$ are separated by two saddle-node bifurcations at $\bar{x}_{\min}$ and $\bar{x}_{\max}$, which are calculated straightforwardly from[69]

$$\frac{\mathrm{d}r(\bar{x})}{\mathrm{d}\bar{x}} \;=\; 0 \quad\Longrightarrow\quad \bar{x}_{\mathrm{crit}}^3\,2(k_1 + l_1) \;-\; \bar{x}_{\mathrm{crit}}^2\,k_1 a_0 \;+\; \kappa k_1 a_0 \;=\; 0\;.$$

As shown in figure 4.45 the integration of the ODE (4.220) reflects precisely the position of S_2.

Stochasticity is readily introduced into the bistable system through Gillespie integration and sampling of trajectories. The results are shown in figure 4.46: At sufficiently small numbers of molecules we observe the system switching back and forth between the two stable states, S_1 and S_3, and an increase in system size changes the scenario in the sense that the system remains in essence in one stable state after it has reached it but identical initial conditions yield either stable state, S_1 or S_3, and the dependence on initial conditions $\mathbf{C}_0$ can only be described probabilistically: $P_{S_1}(\mathbf{C}_0)$ versus $P_{S_3}(\mathbf{C}_0)$. Further increase in system size eventually results in a situation like

[69] In general we obtain three solutions for $\bar{x}_{\mathrm{crit}}$, two of them, $\bar{x}_{\min}$ and $\bar{x}_{\max}$ are situated on the positive $\bar{x}$ axis and correspond to horizontal tangents in the $\bar{x}, r(\bar{x})$-plot (figure 4.45). The corresponding vertical tangents in the $r, \bar{x}(r)$-plot separate the domains with one solution, $0 \le r(\bar{x}) \le r(\bar{x}_{\min})$ and $r(\bar{x}_{\max}) \le r(\bar{x}) < \infty$, and three solutions, $r(\bar{x}_{\min}) \le r(\bar{x}) \le r(\bar{x}_{\max})$.

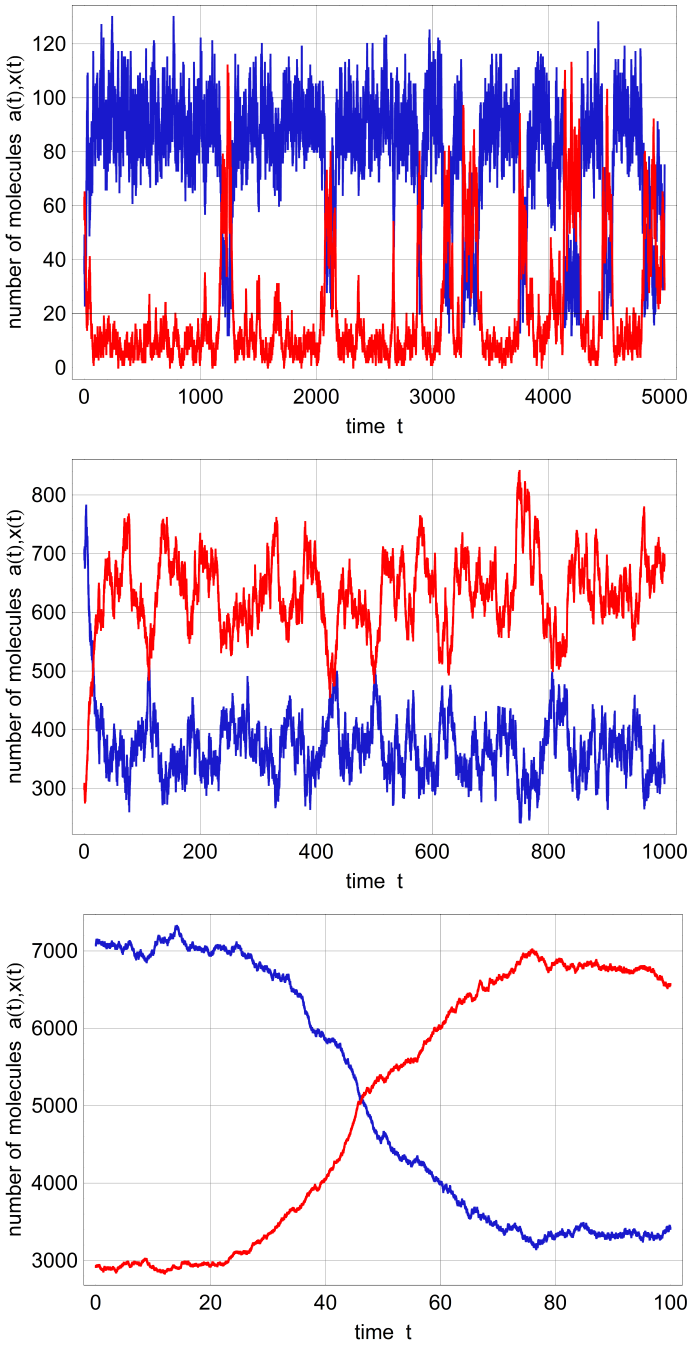


Fig. 4.46 Continued on next page.

Fig. 4.46 Stochasticity in bistable reaction networks. The figure shows three trajectories calculated by means of the Gillespie algorithm with different numbers of molecules: $a_0 = 100$ (top plot), $a_0 = 1000$ (middle plot) and $a_0 = 10\,000$ (bottom plot). For small system sizes a sufficiently long trajectory switches back and forth between the two stable states S_1 and S_3 (top plot). For larger values of a_0 (middle plot) either to S_1 or to S_3 with a ration of the probabilities of approximately 0.56/0.44. At the largest population size (bottom plot) we encounter essentially the same situation as in the deterministic case: the initial conditions determine the state towards which the system converges.

in the deterministic case: Every stable state S_k has its well defined basin of attraction $\mathcal{B}_k$ and in case the initial conditions are situated within the basin, $\mathbf{C}_0 \in \mathcal{B}_k$ the system converges to the attractor S_k. An elegant test for bistability consists in a series of simple experiments: The system in one stable stationary state, S_1 or S_3, is perturbed by additions of increasing quantities of one compound, the system return to the stable state for small perturbations but approaches the other stable state when the perturbation exceeds a certain critical limit. Closely related to this phenomenon is chemical hysteresis that can be easily illustrated by means of figure 4.45: The formation of the stationary state is studied as a function of the flow rate r. Starting from thermodynamic equilibrium at $r = 0$ solution is the the flow reactor state approaches state S_1 until the flow rate $r(\bar{x}_{\max})$ is reached. Then, further increase of r causes the system to jump to state S_3, because S_1 has lost its stability. Alternatively, when the flow rate r is decreased coming from higher values where S_3 is the only stable state, S_3 remains stable until the flow rate $r(\bar{x}_{\min})$ and the solution in the reactor jumps to S_1. Chemical hysteresis implies that different states are passed in the bistable region when going upwards of downwards in the parameter causing bistability.

An experimental reaction mechanism showing bistability is provided by a combination of the Dushman reaction [96],

$$\mathbf{IO_3^-} \; + \; 5\,\mathbf{I^-} \; + \; 6\,\mathbf{H^+} \; \rightleftharpoons \; 3\,\mathbf{I_2} \; + \; 3\,\mathbf{H_2O} \; ,$$

and the Roebuck reaction [385],

$$\mathbf{I_2} \; + \; \mathbf{H_3AsO_3} \; + \; \mathbf{H_2O} \; \rightleftharpoons \; 2\,\mathbf{I^-} \; + \; \mathbf{H_3AsO_4} \; + \; 2\,\mathbf{H^+} \; ,$$

that leads to the overall reaction equation

$$\mathbf{IO_3^-} \; + \; 3\,\mathbf{H_3AsO_3} \; \rightleftharpoons \; \mathbf{I^-} \; + \; 3\,\mathbf{H_3AsO_4} \; ,$$

the oxidation of arsenous acid by iodate. Careful studies of the Dushman-Roebuck reaction in the flow reactor revealed all the features of bistability described here by means of the simpler example [86, 193]. We mention that bistability and hysteresis has also been studied theoretically and experimen-

tally in case of the even more complex Belousov-Zhabotinsky reaction described in the paragraph on the Oregonator model [26, 147, 158]. In all these examples the nonlinear phenomena originate from multistep mechanisms with only monomolecular and bimolecular reaction steps.

Brusselator. The Brusselator mechanism has been invented in Ilya Prigogine's and his group in Brussels [277]. The goal was to find the simplest possible hypothetical chemical system that can sustains oscillations in homogeneous solution and spatial Turing patterns [432] when coupled to diffusion. For this purpose the overall reaction

$$\mathbf{A} + \mathbf{B} \rightleftharpoons \mathbf{D} + \mathbf{E} ,$$

is split into four reaction steps:

$$\mathbf{A} \underset{l_1}{\overset{k_1}{\rightleftharpoons}} \mathbf{X} \tag{4.222a}$$

$$2\,\mathbf{X} + \mathbf{Y} \underset{l_2}{\overset{k_2}{\rightleftharpoons}} 3\,\mathbf{X} \tag{4.222b}$$

$$\mathbf{B} + \mathbf{X} \underset{l_3}{\overset{k_3}{\rightleftharpoons}} \mathbf{Y} + \mathbf{D} \tag{4.222c}$$

$$\mathbf{X} \underset{l_4}{\overset{k_4}{\rightleftharpoons}} \mathbf{E} . \tag{4.222d}$$

As said step (4.222b) is the key to the interesting phenomena of nonlinear dynamics. Compounds **A** and **B** are assumed to be present in buffered concentrations, $[\mathbf{A}] = a_0 = a$ and $[\mathbf{B}] = b_0 = b$, and for the sake of simplicity we consider the case of irreversible reactions, $l_1 = l_2 = l_3 = l_4 = 0$. Then the kinetic differential equations for the deterministic description of the dynamical system are:

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{dt}} &= k_1 a_0 + k_2\, x^2 y - k_3 b_0\, x - k_4\, x \quad \text{and} \\ \frac{\mathrm{d}y}{\mathrm{dt}} &= k_3 b_0\, x - k_2\, x^2 y \ . \end{aligned} \tag{4.223}$$

The Brusselator sustains a single steady state $S = (\bar{x}, \bar{y})$ and conventional bifurcation analysis yields the two eigenvalues $\lambda_{1,2}$. Without loosing generality the analysis is simplified largely by setting all rate constants equal one, $k_1 = k_2 = k_3 = k_4 = 1$:

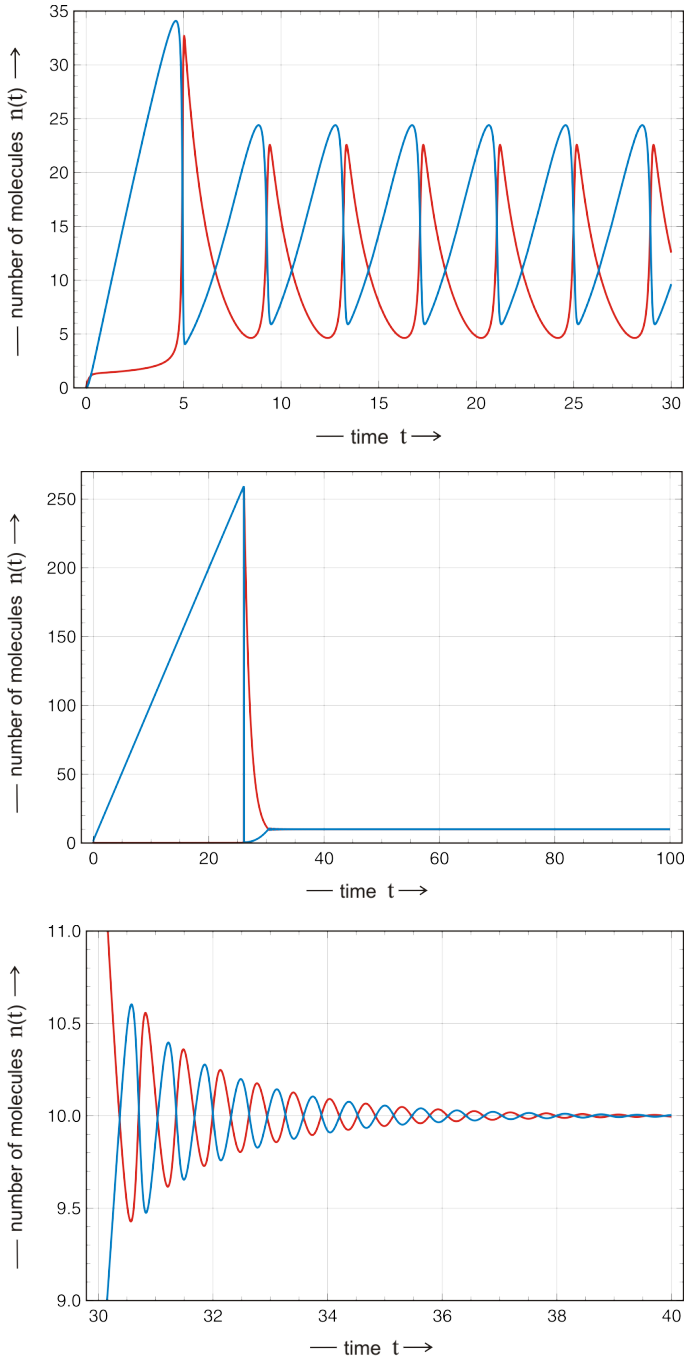


Fig. 4.47 Continued on next page.

Fig. 4.47 Analysis of the Brusselator model. The figure presents integrations of the kinetic differential equation 4.223 in the oscillatory regime (top plot) and in the range of stability of the stationary point $S = (a, b/a)$ (middle plot). Although the integration starts at the origin (0,0), a point that lies relatively close to the stationary state (10,10), the trajectory performs a full refractory cycle before it settles down at the stable point. The plot at the bottom is an enlargement of the middle plot and illustrates the results of a complex conjugate pair of eigenvalues with negative real part: damped oscillations. Parameter choice: $\kappa_1 = k_1 \cdot a_0 = 10$, $\kappa_2 = k_2 = 0.05$, $\kappa_3 = k_3 \cdot b_0 = 6.5$, and $\kappa_4 = k_4 = 1$ (top plot), $\kappa_2 = k_2 = 1$ and $\kappa_3 = k_3 \cdot b_0 = 100$ (middle plot and bottom plot); initial conditions: $x(0) = 0$ and $y(0) = 0$; color code: $x(t)$ red and $y(t)$ blue.

$$S = (\bar{x}, \bar{y}) = \left(a, \frac{b}{a}\right) \quad \text{and} \quad \lambda_{1,2} = \frac{1}{2}\left(b - a^2 - 1 \pm \sqrt{(b - a^2 - 1)^2 - 4a^2}\right),$$

The eigenvalues form a pair complex conjugate values in the parameter range $(a-1)^2 < b < (a+1)^2$ and the real part vanishes at $b = a^2 + 1$. Accordingly, we are dealing with a Hopf bifurcation at $b = a^2 + 1$ with $\lambda_{1,2} = \pm 2\mathrm{i}\, a$. In figure 4.47 we show computer integrations of the ODE (4.223) illustrating the analytical results. For the sake of simplicity we have chosen irreversible reactions and incorporated constant concentrations into the rate constants: $\kappa_1 = k_1 a_0$, $\kappa_2 = k_2$, $\kappa_3 = k_3 b_0$, and $\kappa_4 = k_4$.

Introducing stochasticity into the Brusselator model complicates the bifurcation scenario. At low particle numbers corresponding to a high level of parametric noise the Hopf bifurcation disappears leaving a scenario of more or less irregular oscillations on both sides of the deterministic position of the bifurcation. Ludwig Arnold created the illustrative expression: "Parametric noise destroys the Hopf bifurcation" [17, 18]. Increase in system size allows for the appearance of the stable point attractor on one side of the Hopf bifurcation (figure 4.48).

The oscillations exhibited by the Brusselator are characteristic for so-called excitable media in which a reservoir is filled more or less slowly with a consumable compound until a process rapidly consuming this material is ignited. In case of the Brusselator the consumable is the compound **Y** and its concentration is raised until the autocatalytic process $2\mathbf{X}+\mathbf{Y}\rightarrow 3\mathbf{X}$ is triggered by an above threshold concentration of **Y**. Fast consumption of **Y** results in a rapid increase of **X** that completes the wave by reducing the concentration of **Y** to a small value (figure 4.49). The easiest way to visualize an excitable medium is provided by the example of wildfires: Wood grows slowly until it reaches a density that can sustain spreading fire. After triggered by natural causes or arson the fire consumes all the wood and thereby initiates the basis for the next refractory period. Oscillatory chemical reactions do not need an external trigger since an internal fluctuation is sufficient to initiate the decay phase.

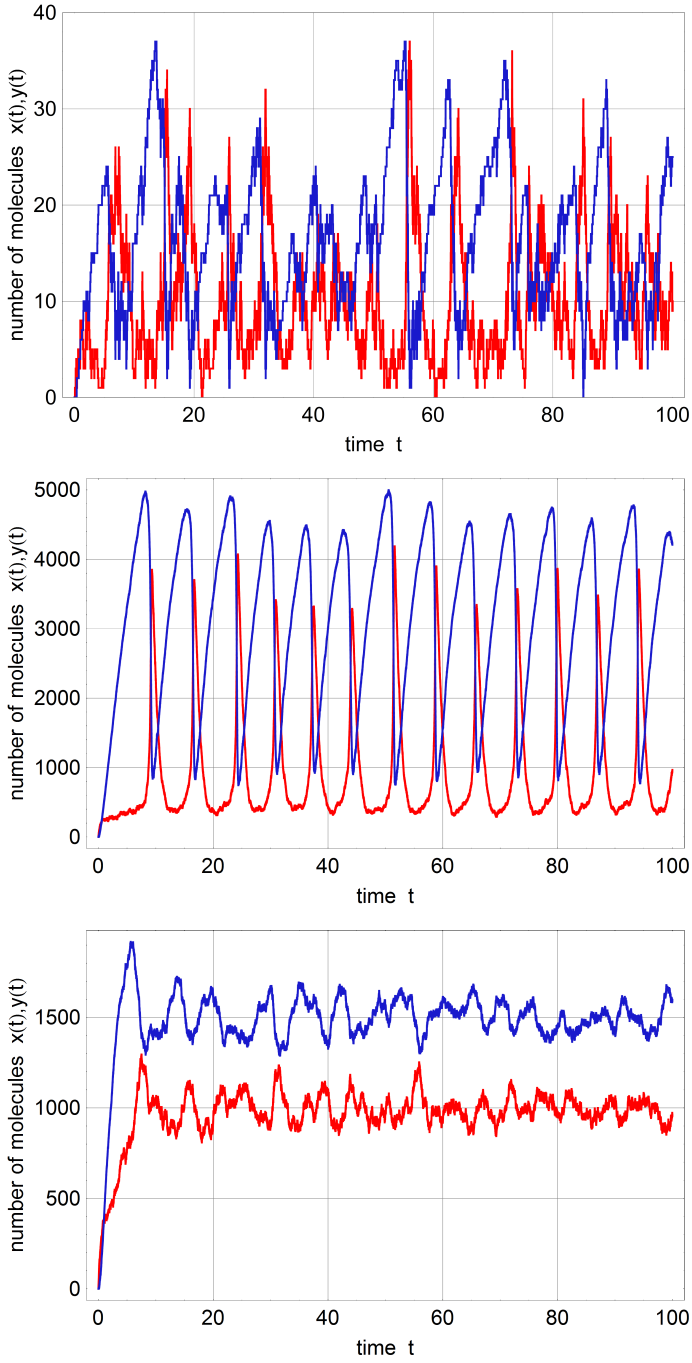


Fig. 4.48 Continued on next page.

Fig. 4.48 Stochasticity in the Brusselator model. The figure shows three stochastic simulations of the Brusselator model. The top plot shows the Brusselator in the stable regime at low numbers of molecules ($a_0 = 10$) but no settling down of the trajectories near the steady state can be observed. At sufficiently high numbers of molecules ($a_0 = 1000$) the behavior of the stochastic Brusselator is close to the deterministic solutions (figure 4.47) in the oscillatory regime (middle plot) and in the range of fixed point stability the stochastic solutions fluctuate around the stationary values (bottom plot). Parameter choice: $\kappa_1 = 10, \kappa_2 = 0.01, \kappa_3 = 1.5, \kappa_4 = 1$ (top plot), $\kappa_1 = 1000, \kappa_2 = 1 \times 10^{-6}, \kappa_3 = 3, \kappa_4 = 1$ (middle plot), and $\kappa_1 = 1000, \kappa_2 = 1 \times 10^{-6}, \kappa_3 = 1.5, \kappa_4 = 1$ (bottom plot); iniital conditions: $x(0) = y(0) = 0$; color code: x red and y blue.

Oregonator. The prototype of an oscillatory chemical reaction and the first example that became popular in history is the Belousov-Zhabotinsky reaction, which is described by the overall process of the cerium catalyzed reaction of malonic acid by bromate ions in dilute sulfuric acid. We mention the reaction here in order to present one example showing how complicated chemical reaction networks can be in reality:

$$3\,\mathbf{H_2M} + 4\,\mathbf{BrO_3^-} \longrightarrow 4\,\mathbf{Br^-} + 9\,\mathbf{CO_2} + 6\,\mathbf{H_2O}\ . \tag{4.224a}$$

Malonic acid is written here as $\mathbf{CH_2(CO_2H)_2} \equiv \mathbf{H_2M}$. The process can be split into three overall reactions that follow the reaction equations

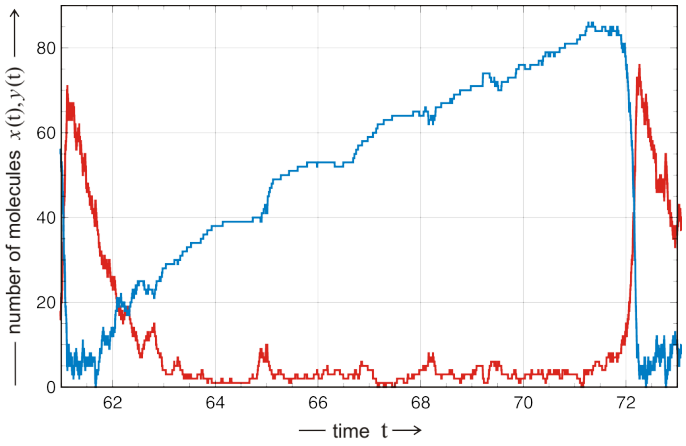


Fig. 4.49 Refractory cycle in the Brusselator model. The figure presents an enlargement from a stochastic trajectory calculated with the parameters applied in the top plot of figure 4.48. It illustrates a refractory cycle consisting of filling a reservoir of compound **Y** (blue) that is quickly emptied by conversion of **Y** to **X** after ignition triggered by a sufficiently large concentration of **Y**.

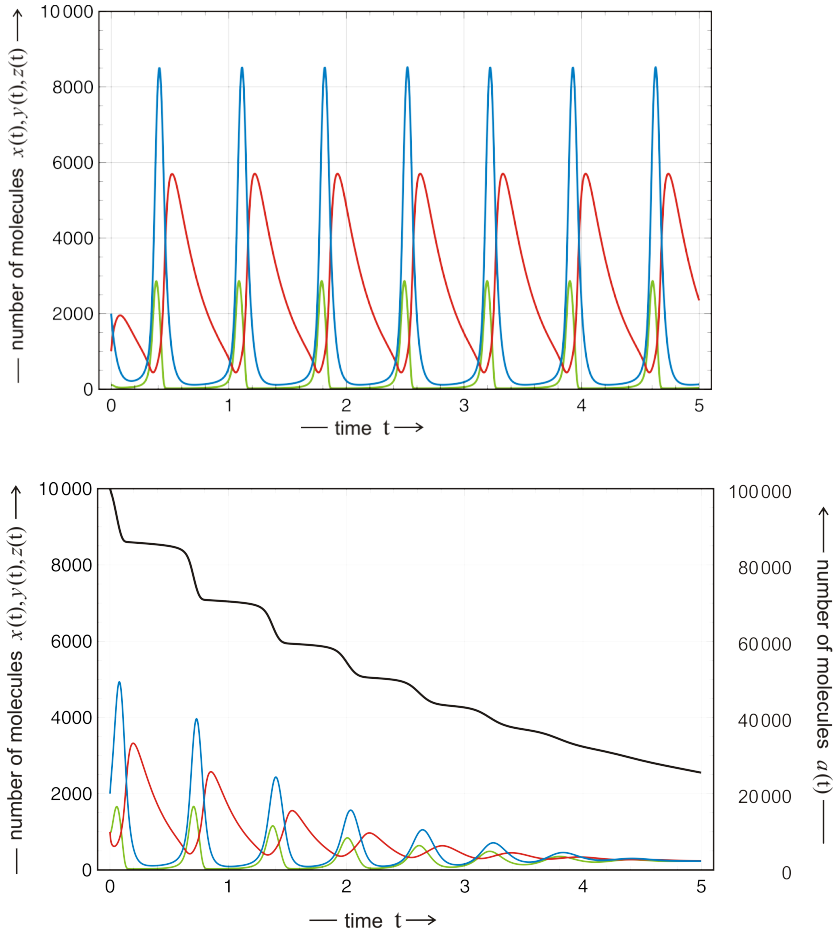


Fig. 4.50 Analysis of the Oregonator model. The kinetic ODEs of the Oregonator model (4.225f) are integrated and the undamped oscillations in the open system (with buffered concentrations of **A** and **B**) are shown in the top plot. The supply of **A** is limited in the bottom plot that mimics the closed system. As **A** is consumed the oscillations become smaller an eventually die out. Parameter choice: $\kappa_1 = 2, \kappa_2 = 0.1, \kappa_3 = 104, \kappa_4 = 0.016$, and $\kappa_5 = 26$: initial concentrations $x(0) = 100$, $y(0) = 1000$, and $z(0) = 2000$; color code: $x(t)$ green, $y(t)$ red, $z(t)$ blue, and $a(t)$ black.

$$2\,\mathbf{Br}^- + \mathbf{BrO}_3^- + 3\,\mathbf{H}^+ + 3\,\mathbf{H_2M} \longrightarrow 3\,\mathbf{HBrM} + 3\,\mathbf{H_2O}\,, \tag{4.224b}$$

$$4\,\mathbf{Ce}^{3+} + \mathbf{BrO}_3^- + 5\,\mathbf{H}^+ \longrightarrow 4\,\mathbf{Ce}^{4+} + \mathbf{HBrO} + 2\,\mathbf{H_2O}\,, \quad \text{and} \tag{4.224c}$$

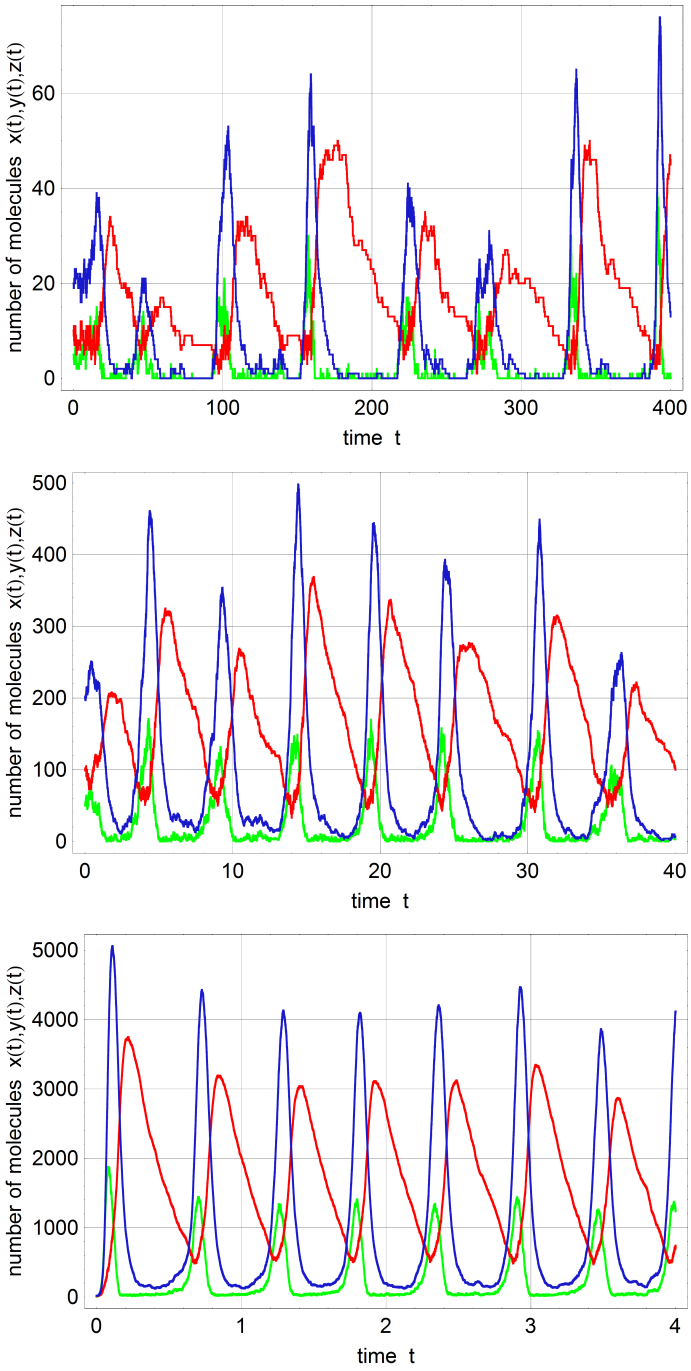


Fig. 4.51 Continued on next page.

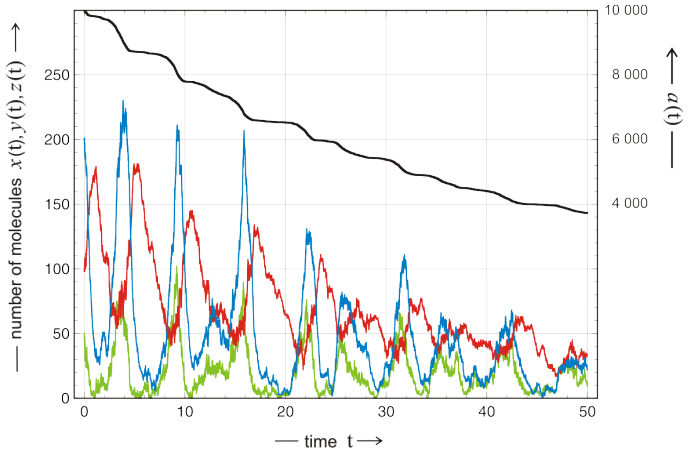


Fig. 4.51 Stochasticity in the Oregonator model. The figure shows stochastic simulations of the Oregonator model at different population sizes: $x(0) = 5 \cdot \vartheta$, $y(0) = 10 \cdot \vartheta$ (red), and $z(0) = 20 \cdot \vartheta$ with $\vartheta = 1$ (previous page, top plot), $\vartheta = 10$ (previous page, middle plot), and $\vartheta = 100$ (previous page, bottom plot). A simulation of the Oregonator in a system that is closed with respect to compound **A** is shown on this page ($\vartheta = 10$, $a(0) = 10\,000$). The parametrization was adopted from [167]: $\bar{x} = 5$, $\bar{y} = 10$, and $\bar{z} = 20$ were used for the concentrations, and $\bar{\rho}_1 = 0.2 \cdot \vartheta^2$ and $\bar{\rho}_2 = 5 \cdot \vartheta^2$ for the reaction rates at the unstable stationary point, and this yields for the rate parameters, $\kappa_1 = 0.02\vartheta$, $\kappa_2 = 0.1$, $\kappa_3 = 1.04\vartheta$, $\kappa_4 = 0.016\vartheta$, and $\kappa_5 = 0.26\vartheta$. Color code: $x(t)$ green, $y(t)$ red, $z(t)$ blue, and $a(t)$ black.

$$\begin{aligned} 2\,\mathbf{Ce}^{4+} + 2\,\mathbf{H_2M} + \mathbf{HBrM} + \mathbf{HBrO} + 2\,\mathbf{H_2O} \longrightarrow \\ \longrightarrow 2\,\mathbf{Ce}^{3+} + 2\,\mathbf{Br}^{-} + 3\,\mathbf{HOHM} + 4\,\mathbf{H}^{+}\,. \end{aligned} \tag{4.224d}$$

The third reaction (4.224d) is complemented by the fragmentation of hydroximalonic acid to carbon dioxide and accompanied by further reduction of bromate to bromide. In detail the reaction mechanism is even more complicated and a network of about 20 reaction steps has been derived from the available data [99].

For a complete theoretical analysis the reaction network of the Belousov-Zhabotinsky reaction in too complicated and therefore a simplified model, the Oregonator model has been conceived by US American physical chemists Richard Noyes and Richard Field in order to allow for a combined analytical and numerical study of a system that mimics closely the properties of the Belousov-Zhabotinsky reaction [137, 350]:

$$\mathbf{A} + \mathbf{Y} \underset{l_1}{\overset{k_1}{\rightleftharpoons}} \mathbf{X} + \mathbf{P} \tag{4.225a}$$

$$\mathbf{X} + \mathbf{Y} \underset{l_2}{\overset{k_2}{\rightleftharpoons}} 2\,\mathbf{P} \tag{4.225b}$$

$$\mathbf{A} + \mathbf{X} \underset{l_3}{\overset{k_3}{\rightleftharpoons}} 2\,\mathbf{X} + 2\,\mathbf{Z} \tag{4.225c}$$

$$2\,\mathbf{X} \underset{l_4}{\overset{k_4}{\rightleftharpoons}} \mathbf{A} + \mathbf{B} \tag{4.225d}$$

$$\mathbf{B} + \mathbf{Z} \xrightarrow{k_5} \frac{1}{2} f\,\mathbf{Y}\,. \tag{4.225e}$$

The corresponding kinetic ODEs for irreversible reactions and buffered concentration of **A** and **B** with $[\mathbf{X}] = x$, $[\mathbf{Y}] = y$, and $[\mathbf{Z}] = z$ are:

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= k_1 a\, y - k_2\, xy + k_3 a\, x - 2k_4\, x^2\,, \\ \frac{\mathrm{d}y}{\mathrm{d}t} &= -k_1 a\, y - k_2\, xy + \frac{1}{2} k_5 b\, z\,, \quad \text{and} \\ \frac{\mathrm{d}z}{\mathrm{d}t} &= 2k_3 a\, x - k_5 b\, z\,. \end{aligned} \tag{4.225f}$$

Two features of the model are remarkable: (i) It is low-dimensional – three variables $[\mathbf{X}]$, $[\mathbf{Y}]$, and $[\mathbf{Z}]$ when **A** and **B** are buffered – and does not contain a termolecular step, and (ii) it makes use of a non-stoichiometric factor f. The Oregonator model has been successfully applied to reproduce experimental findings on fine details of the oscillations in the Belousov-Zhabotinsky reaction but failed to predict the occurrence of deterministic chaos. In later works new models on this reaction have been developed that were successful also in this aspect [152, 185].

In figure 4.50 we show numerical integrations of equation (4.225f) in the open system with constant input materials and in a closed system with limited supply of **A**. Interestingly the oscillations give rise to a step consumption of the resource. In his seminal paper on the simulation algorithm and its applications Daniel Gillespie [167] provided a stochastic version of the Oregonator, which we are applying here to demonstrate the approach of the stochastic simulations towards the deterministic solution with increasing population size (figure 4.51).

References

1. Abramowitz, M., Segun, I.A. (eds.): Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications, New York (1965)
2. Abramson, M., Moser, W.O.J.: More birthday surprises. Amer. Math. Monthly **77**, 856–858 (1970)
3. Acton, F.S.: Numerical Methods that Work. Harper & Row, Publishers, New York (1970)
4. Acton, F.S.: Numerical Methods that (usually) Work, fourth printing edn. Mathematical Association of America, Washington, DC (1990)
5. Adams, W.J.: The Life and Times of the Central Limit Theorem, *History of Matheatics*, vol. 35, second edn. American Mathematical Society and London Mathematical Society, Providence, RI (2009). Articles by A. M. Lyapunov translated from the Russian by Hal McFaden.
6. Al-Soufi, W., Reija, B., Novo, M., Kelekyan, S., Kühnemuth, R., Seidel, C.A.M.: Fluorescence correlation sprctroscopy, a tool to inverstigate supramolecular dynamics: Inclusion complexes of pyronines with cyclodextrin. J. Am. Chem. Soc. **127**, 8775–8784 (2005)
7. Anderson, B.D.O.: Reverse-time diffusion equation models. Stochastic Processes and Applications **12**, 313–326 (1982)
8. Anderson, D.F.: Incorporating postleap checks in tau-leaping. J. Chem. Phys. **128**, e 054103 (2008)
9. Anderson, D.F., Craciun, G., Kurtz, T.G.: Product-form stationary distributions for deficiency zero chemical reaction networks. Bull. Math. Biol. **72**, 1947–1970 (2010)
10. Anderson, D.F., Ganguly, A., Kurtz, T.G.: Error analysis of tau-leap simulation methods. Annals Appl. Probability **6**, 2226–2262 (2011)
11. Applebaum, D.: Lévy processes – From probability to finance and quantum groups. Notices Am. Math. Soc. **51**, 1336–1347 (2004)
12. Aragón, S.R., Pecora, R.: Fluorescence correlation spectroscopy and Brownian rotational diffusion. Biopolymers **14**, 119–138 (1975)
13. Arányi, P., Tóth, J.: A full stochastic description of the Michaelis-Menten reaction for small systems. Acta Biochim. et Biophys. Acad. Sci. Hung. **12**, 375–388 (1977)
14. Arfken, G.B., Weber, H.J.: Mathematical Methods for Physicists, fifth edn. Harcourt Academic Press, San Diego, CA (2001)
15. arlin, S.K., McGregor, J.: On a genetics model of moran. Math. Proc. Camb. Phil. Soc. **58**, 299–311 (1962)

16. Arnold, L.: Stochastic Differential Equations. Theory and Applications. John Wiley & Sons, New York (1974)
17. Arnold, L.: Random Dynamical Systems. Springer-Verlag, Berlin (1998). Second corrected printing 2003
18. Arnold, L., Bleckert, G., Schenk-Hoppé, K.R.: The stochastic brusselator: Parametric noise destroys hopf bifurcation. In: H. Crauel, M. Gundlach (eds.) Stochastic Dynamics, chap. 4, pp. 71–92. Springer, New York (1999)
19. Arscott, F.M.: Heun's equation. In: A. Ronveau (ed.) Heun's Differential Equations, pp. 3–86. Oxford University Press, New York (1955)
20. Arslan, E., Laurenzi, I.J.: Kinetics of autocatalysis in small systems. J. Chem. Phys. **128**, e 015101 (2008)
21. Aster, R.C., Borchers, B., Thurber, C.H.: Parameter Estimation and Inverse Problems, second edn. Academic Press, Elsevier, Singapore (2013)
22. Athreya, K.B., Ney, P.E.: Branching Processes. Springer-Verlag, Heidelberg, DE (1972)
23. Atkins, P.W., Friedman, R.S. (eds.): Molecular Quantum Mechanics, fifth edn. Oxford University Press, Oxford, UK (2010)
24. Bachelier, L.: Théorie de la spéculation. Annales scientifiques de l'É.N.S. 3^e série **17**, 21–86 (1900)
25. Bailey, N.T.J.: The Elements of Stochastic Processes with Application in the Natural Sciences. Wiley, New York (1964)
26. Bar-Eli, K., Noyes, R.M.: Detailed calculations of multiple steady states during oxidation of cerous ion by bromate in a stirred flow reactor. J. Phys. Chem. **82**, 1352–1359 (1978)
27. Bartholomay, A.F.: On the linear birth and death processes of biology as Markoff chains. Bull. Math. Biophys. **20**, 97–118 (1958)
28. Bartholomay, A.F.: Stochastic models for chemical reactions: I. Theory of the unimolecular reaction process. Bull. Math. Biophys. **20**, 175–190 (1958)
29. Bartholomay, A.F.: Stochastic models for chemical reactions: II. The unimolecular rate constant. Bull. Math. Biophys. **21**, 363–373 (1959)
30. Bartholomay, A.F.: A stochastic approach to statistical kinetics with applications to enzyme kinetics. Biochemistry **1**, 223–230 (1962)
31. Bartlett, M.S.: Stochastic processes or the statistics of change. J. Roy. Statist. Soc. C **2**, 44–64 (1953)
32. Bazley, N.W., Montroll, E.W., Rubin, R.J., Shuler, K.E.: Studies in nonequilibrium rate processes: III. The vibrational relaxation of a system of anharmonic oscillators. J. Chem. Phys. **28**, 700–704 (1958). Erratum: *J.Chem.Phys.*, 29:1185-1186
33. Berg, J.M., Tymoczko, J.L., Stryer, L.: Biochemistry, fifth edn. W. H. Freeman and Company, New York (2002)
34. Berg, J.M., Tymoczko, J.L., Stryer, L.: Biochemistry, seventh edn. W. H. Freeman and Company, New York (2012)
35. Bergström, H.: On some expansions of stable distribution functions. Ark. Math. **2**, 375–378 (1952)
36. Berry, R.S., Rice, S.A., Ross, J.: Physical Chemistry, second edn. Oxford University Press, New York (2000)
37. Berry, R.S., Rice, S.A., Ross, J.: Physical and Chemical Kinetics, second edn. Oxford University Press, New York (2002)
38. Biebricher, C.K., Eigen, M., William C. Gardiner, J.: Kinetics of RNA replication. Biochemistry **22**, 2544–2559 (1983)
39. Bienaymé, I.J.: Da la loi de Multiplication et de la durée des familles. Soc. Philomath. Paris Extraits **Ser. 5**, 37–39 (1845)
40. Billingsley, P.: Probability and Measure, third edn. Wiley-Interscience, New York (1995)

41. Billingsley, P.: Probability and Measure, Anniversary edn. Wiley-Interscience, Hoboken, NJ (2012)
42. Binnig, G., Quate, C.F., Gerber, C.: Atomic force microscopy. Phys. Rev. Lett. **56**, 930–933 (1986)
43. Björck, Å.: Numerical Methods for Least Square Problems. Other Titles in Applied Mathematics. SIAM Society for Industrial & Applied Mathematics, Philadelphia, PA (1996)
44. Bloomfield, V.A., Benbasat, J.A.: Inelastic light-scattering study of macromolecular reaction kinetics. I: The reactions $A \rightleftharpoons B$ and $2A \rightleftharpoons A_2$. Macromolecules **4**, 609–613 (1971)
45. Blythe, R.A., McKane, A.J.: Stochastic models of evolution in genetics, ecology and linguistics. J. Stat. Mech,: Theor. Exp. (2007). P07018
46. Boas, M.L.: Mathematical Methods in the Physical Sciences, third edn. John Wiley & Sons, Hoboken, NJ (2006)
47. Boole, G.: An Investigation of the Laws of Thought on which are Founded the Mathematical Theories of Logic and Probabilities. MacMillan, London (1854). Reprinted by Dover Publ. Co., New York, 1958
48. Born, M., Oppenheimer, R.: Zur Quantentheorie der Moleküle. Annalen der Physik **84**, 457–484 (1927). In German
49. Bouchaud, J.P., Georges, A.: Anomalous diffusion in disordered madia: Statistical mechanisms, models and physical applications. Physics Reports **195**, 127–293 (1990)
50. Box, G.E.P., Muller, M.E.: A note on the generation of random normal deviates. Annals Math. Statist. **29**, 610–611 (1958)
51. Brenner, S.: Theoretical biology in the third millenium. Phil. Trans. Roy. Soc. London B **354**, 1963–1965 (1999)
52. Brenner, S.: Hunters and gatherers. The Scientist **16**(4), 14 (2002)
53. Brockwell, P.J., Davis, R.A.: Introduction to Time Series and Forecasting. Springer-Verlag, New York (1996)
54. Brockwell, P.J., Davis, R.A., Yang, Y.: Continuous-time Gaussian autoregression. Statistica Sinica **17**, 63–80 (2007)
55. Brown, R.: A brief description of microscopical observations made in the months of June, July and August 1827, on the particles contained in the pollen of plants, and on the general existence of active molecules in organic and inorganic bodies. Phil. Mag., Series 2 **4**, 161–173 (1828). First Publication: The Edinburgh New Philosophical Journal. July-September 1828, pp.358-371.
56. Calaprice, A. (ed.): The Ultimate Quotable Einstein. Princeton University Press, Princeton, NJ (2010)
57. de Candolle, A.: Zur Geschichte der Wissenschaften und Gelehrten seit zwei Jahrhunderten nebst anderen Studien über wissenschaftliche Gegenstände insbesondere über Vererbung und Selektion beim Menschen. Akademische Verlagsgesellschaft, Leipzig, DE (1921). Deutsche Übersetzung der Originalausgabe *"Histoire des sciences et des savants depuis deux siècle"*, Geneve 1873, durch Wilhelm Ostwald.
58. Cann, R.L.: Y weigh in again on modern humans. Science **341**, 465–467 (2013)
59. Cann, R.L., Stoneking, M., Wilson, A.C.: Mitochondrial DNA and human evolution. Nature **325**, 31–36 (1987)
60. Cao, Y., Gillespie, D.T., Petzold, L.R.: Efficient step size selection for the tau-leaping simulation method. J. Chem. Phys. **124**, 044,109 (2004)
61. Cao, Y., Gillespie, D.T., Petzold, L.R.: Avoiding negative populations in explicit Poisson tau-leaping. J. Chem. Phys. **123**, e054,104 (2005)
62. Cao, Y., Gillespie, D.T., Petzold, L.R.: Efficient step size selection for the tau-leaping simulation method. J. Chem. Phys. **124**, e044,109 (2006)
63. Cao, Y., Gillespie, D.T., Petzold, L.R.: Adaptive explicit-implicit tau-leaping method with automatic tau selection. J. Chem. Phys. **126**, e224,101 (2007)

64. Carter, M., van Brunt, B.: The Lebesgue-Stieltjes Integral. A Practical Introduction. Springer-Verlag, Berlin (2007)
65. Chang, C., Gzyl, H.: Parameter estimation in superposition of decaying exponentials. Applied Mathemtics and Computation **96**, 101–116 (1998)
66. Chechkin, A.V., Metzler, R., Klafter, J., Gonchar, V.Y.: Introduction to the theory of Lévy flights. In: R. Klages, G. Radons, I.M. Sokolov (eds.) Anomalous Transport: Foundations and Applications, chap. 5, pp. 129–162. Wiley-VCH Verlag GmbH, Weinheim, DE (2008)
67. Child, M.S.: Molecular Collision Theory. Dover Publications, Mineola, NY (1996). Originally publisher: Academic Press, London 1974.
68. Chung, K.L.: A Course in Probability Theory, *Probability and Mathematical Statistics*, vol. 21, second edn. Academic Press, New York (1974)
69. Chung, K.L.: Elementary Probability Theory with Stochastic Processes, 3rd edn. Springer-Verlag, New York (1979)
70. Cochran, W.G.: The distribution of quadratic forms in normal systems, with applications to the analysis of covariance. Math. Proc. Cambridge Phil. Soc. **30**, 178–191 (1934)
71. Conrad, K.: Probability distributions and maximum entropy. Expository paper, University of Connecticut, Storrs, CT (2005)
72. Cook, M., Soloveichik, D., Winfree, E., Bruck, J.: Programmability of chemical reaction networks. In: A. Condon, D. Harel, J.N. Kok, A. Salomaa, E. Winfree (eds.) Algorithimc Bioprocesses, *Natural Computing Series*, vol. XX, pp. 543–584. Springer-Verlag, Berlin (2009)
73. Cooper, B.E.: Statistics for Experimentalists. Pergamon Press, Oxford (1969)
74. Cortina Borja, M., Haigh, J.: The birthday problem. Significance **4**, 124–127 (2007)
75. Cover, T.M., Thomas, J.A.: Elements of Information Theory, second edn. John Wiley & Sons, Hoboken, NJ (2006)
76. Cox, D.R., Miller, H.D.: The Theory of Stochastic Processes. Methuen, London (1965)
77. Cox, R.T.: The Algebra of Probable Inference. The John Hopkins Press, Baltimore, MD (1961)
78. Craciun, G., Tang, Y., Feinberg, M.: Understanding bistability in complex enzyme-driven reaction networks. Proc. Natl. Acad. Sci. USA **103**, 8697–8702 (2006)
79. Crank, J.: The Mathematics of Diffusion. Clarendon Press, Oxford, UK (1956)
80. Crow, J.F., Kimura, M.: An Introduction to Population Genetics Theory. Sinauer Associates, Sunderland, MA (1970). Reprinted at *The Blackburn Press*, Caldwell, NJ, 2009
81. Cull, P., Flahive, M., Robson, R.: Difference Equations. From Rabbits to Chaos. Undergraduate Texts in Mathematics. Springer, New York (2005)
82. Dalla Valle, J.M.: Note on the Heaviside expansion formula. Proc. Natl. Acad. Sci. USA **17**, 678–684 (1931)
83. Darvey, I.G., Ninham, B.W.: Stochastic models for second-order chemical reaction kinetics. Time course of reactions. J. Chem. Phys. **46**, 1626–1645 (1967)
84. Darvey, I.G., Ninham, B.W., Staff, P.J.: Stochastic models for second-order chemical reaction kinetics. The equilibirum state. J. Chem. Phys. **45**, 2145–2155 (1966)
85. Darvey, I.G., Staff, P.J.: Stochastic approach to first-order chemical reaction kinetics. J. Chem. Phys. **44**, 990–997 (1966)
86. DeKepper, P., Epstein, I.R., Kustin, K.: Bistability in the oxidatiion of arsenite by iodate in a stirred flow reactor. J. Am. Chem. Soc. **103**, 6121–6127 (1981)
87. Delbrück, M.: Statistical fluctuations in autocatalytic reactions. J. Chem. Phys. **8**, 120–124 (1940)

88. Demetrius, L., Schuster, P., Sigmund, K.: Polynucleotide evolution and branching processes. Bull. Math. Biol. **47**, 239–262 (1985)
89. Devroye, L.: Non-Uniform Random Variate Generation. Springer-Verlag, New York (1986)
90. Djermoune, E.H., Tomczak, M.: Statistical analysis of the Kumaresan-Tufts and matrix pencil methods in estimating damped sinusoids. In: F. Hlawatsch, G. Matz, M. Rupp, B. Wistawel (eds.) Proceedings of the XII. European Signal Processing Conference, vol. II, pp. 1261–1264. Technische Universität Wien, Wien (2004)
91. Domingo, E., Parrish, C.R., John J, H. (eds.): Origin and Evolution of Viruses, second edn. Elsevier, Academic Press, Amsterdam, NL (2008)
92. Donnelly, P.J., Tavaré, S.: Coalescents and genealogical structure under neutrality. Annu. Rev. Genet. **29**, 401–421 (1995)
93. Doob, J.L.: Topics in the theory of Markoff chains. Trans. Am. Math. Soc. **52**, 37–64 (1942)
94. Doob, J.L.: Markoff chains – Denumerable case. Trans. Am. Math. Soc. **58**, 455–473 (1945)
95. Dudley, R.M.: Real Analysis and Probability. Wadsworth and Brooks, Pacific Grove, CA (1989)
96. Dushman, S.: The reaction between iodic and hydroiodic acid. J. Phys. Chem. **8**, 453–482 (1903)
97. Dyson, F.: A meeting with Enrico Fermi. How one intuitive physicist rescued a team from fruitless research. Nature **427**, 297 (2004)
98. Eddy, S.R.: What is Bayesian statistics? Nature Biotechnology **22**, 1177–1178 (2004)
99. Edelson, D., Field, R.J., Noyes, R.M.: Mechanistic details of the Belousov-Zhabotinskii oscillations. Internat. J. Chem. Kinetics **7**, 417–423 (1975)
100. Edman, L., Földes-Papp, Z., Wennmalm, S., Rigler, R.: The fluctuating enzyme: A single moleculae approach. Chem. Phys. **247**, 11–22 (1999)
101. Edman, L., Rigler, R.: Memory landscapes of single-enzyme molecules. Proc. Natl. Acad. Sci. USA **97**, 8266–8271 (2000)
102. Edwards, A.W.F.: Are Mendel's resulta really too close˙ Biological Reviews **61**, 295–312 (1986)
103. Ehrenberg, M., Rigler, R.: Rotational Brownian motion and fluorescence intensity fluctuations. Chem. Phys. **4**, 390–401 (1974)
104. Ehrenfest, P., Ehrenfest, T.: Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem. Z. Phys. **8**, 311–314 (1907)
105. Eigen, M.: Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften **58**, 465–523 (1971)
106. Eigen, M., McCaskill, J., Schuster, P.: The molecular quasispecies. Adv. Chem. Phys. **75**, 149–263 (1989)
107. Eigen, M., Schuster, P.: The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. Naturwissenschaften **64**, 541–565 (1977)
108. Einstein, A.: Über die von der molekular-kinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. Annal. Phys. (Leipzig) **17**, 549–560 (1905)
109. Einstein, A.: Investigations on the Theory of the Brownian Movement. Dover Publications, New York (1956). Five original publications by Albert Einstein edited with notes by R. Fürth
110. Elliot, R.J., Anderson, B.D.O.: Reverse-time diffusions. Stochastic Processes and Applications **19**, 327–339 (1985)
111. Elliot, R.J., Kopp, A.E.: Mathematics of Financial Markets, second edn. Springer-Verlag, New York (2005)

112. Elson, E., Magde, D.: Fluorescence correlation spectroscopy. I. Conceptual basis and theory. Biopolymers **13**, 1–27 (1974)
113. Engl, H.W., Flamm, C., Kügler, P., Lu, J., Müller, S., Schuster, P.: Inverse problems in systems biology. Inverse Problems **25**, 123,014 (2009)
114. Engl, H.W., Hanke, M., Neubauuer, A.: Regularization of Inverse Problems. Kluwer Academic Publishers, Boston, MA (1996)
115. Érdi, P., Lente, G.: Stochastic Chemical Kinetics. Theory and (mostly) Systems Biological Applications. Understanding Complex Systems. Springer Verlag, Berlin (2014)
116. Evans, M., Hastings, N.A.J., Peacock, J.B.: Statistical Distributions, third edn. John Wiley & Sons, New York (2000)
117. Everett, C.J., Ulam, S.: Multiplicative systems I. Proc. Natl. Acad. Sci. USA **34**, 403–405 (1948)
118. Everett, C.J., Ulam, S.M.: Multiplicative systems in several variables I. Tech. Rep. LA-683, Los Alamos Scientific Laboratory (1948)
119. Everett, C.J., Ulam, S.M.: Multiplicative systems in several variables II. Tech. Rep. LA-690, Los Alamos Scientific Laboratory (1948)
120. Everett, C.J., Ulam, S.M.: Multiplicative systems in several variables III. Tech. Rep. LA-707, Los Alamos Scientific Laboratory (1948)
121. Ewens, W.J.: Mathematical Population Genetics. I. Theoretical Introduction, second edn. Interdisciplinary Applied Mathematics. Springer-Verlag, Berlin (2004)
122. Eyring, H.: The activated complex in chemical reactions. J. Chem. Phys. **3**, 107–115 (1935)
123. Farlow, S.J.: Partial Differential Equations for Scientists and Engineers. Dover Publications, New York (1982)
124. Feinberg, M.: Complex balancing in general kinetic systems. Archive for Rational Mechanics and Analysis **49**, 187–194 (1972)
125. Feinberg, M.: Mathematical aspects of mass action kinetics. In: L. Lapidus, N.R. Amundson (eds.) Chemical Reactor Theory – A Review, pp. 1–78. Prentice Hall, Englewood Cliffs, NJ (1977)
126. Feinberg, M.: Lectures on Chemical Reaction Networks. Chemical Engineering & Mathematics. The Ohio State University, Columbus, OH (1979)
127. Feinberg, M.: Chemical oscillations, multiple equilibria, and reaction network structure. In: W.E. Stewart, W.H. Ray, C.C. Conley (eds.) Dynamics and Modelling of Reactive Systems, pp. 59–130. Academic Press, New York (1980)
128. Feinberg, M.: Chemical reaction network structure and the stability of complex isothermal reactors – II. Multiple steady states for networks of deficiency one. Chemical Engineering Science **43**, 1–25 (1988)
129. Feller, W.: On the integro-differential equations of purely discontinuous Markoff processes. Trans. Amer. Math. Soc. **48**, 488–515 (1940)
130. Feller, W.: The general form of the so-called law of the iterated logarithm. Trans. Amer. Math. Soc. **54**, 373–402 (1943)
131. Feller, W.: On the theory of stochastic processes, with particular reference to applications. In: The Regents of the University of California (ed.) Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, pp. 403–432. University of California Press, Berkeley, CA (1949)
132. Feller, W.: Diffusion processes in genetics. In: J. Neyman (ed.) Proc. 2nd Berkeley Symp. on Mathematical Statistics and Probability. University of Caifornia Press, Berkeley, CA (1951)
133. Feller, W.: An Introduction to Probability Theory and Its Application, vol. I, third edn. John Wiley & Sons, New York (1968)
134. Feller, W.: An Introduction to Probability Theory and Its Application, vol. II, second edn. John Wiley & Sons, New York (1971)

135. Fersht, A.: Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. W. H. Fremman and Company, New York (1999)
136. Fick, A.: Über diffusion. Annalen der Physik und Chemie **170 (4. Reihe 94)**, 59–86 (1855)
137. Field, R.J., Körös, E., Noyes, R.M.: Oscillations in chemical systems. II. Thorough analysis of temporal oscillations in the bromate-cerium-malonic acid system. J. Am. Chem. Soc. **94**, 8649–8664 (1972)
138. Field, R.J., Noyes, R.M.: Oscillations in chemical systems. IV. Limit cycle behavior in a model of a real chemical reaction. J. Chem. Phys. **60**, 1877–1884 (1974)
139. Fisher, R.A.: Applications of "Student's" distribution. Metron **5**, 90–104 (1925)
140. Fisher, R.A.: Moments and product moments of sampling distributions. Proc. London Math. Soc. **Ser.2, 30**, 199–238 (1928)
141. Fisher, R.A.: The Genetical Theory of Natural Selection. Oxford University Press, Oxford, UK (1930)
142. Fisher, R.A.: Has Mendel's work been rediscovered? Annals of Science pp. 115–137 (1936)
143. Fisher, R.A.: The Design of Experiments, eighth edn. Hafner Publishing Company, Edinburgh, UK (1966)
144. Fisk, D.L.: Quasi-martingales. Trans. Amer. Math. Soc. **120**, 369–389 (1965)
145. Fisz, M.: Probability Theory and Mathematical Statistics, third edn. John Wiley & Sons, New York (1963)
146. Fisz, M.: Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaft, Berlin (1989). In German
147. Föllner, H.H., Geiseler, W.: A model of bistability in an open homogeneous chemical reaction system. Naturwissenschaften **64**, 384 (1977)
148. Foster, D.P.: Law of the iterated logarithm. Wikipedia entry, University of Pennsylvania, Philadelphia, PA (2009). Retrieved April 07, 2009 from `en.wikipedia.org/wiki/Law_of_the_iterated_logarithm`
149. Francalacci, P., Morelli, L., Angius, A., Berutti, R., Reinier, F., Atzeni, R., Pilu, R., Busonero, F., Maschino, A., Zara, I., Sanna, D., Useli, A., Urru, M.F., Marcelli, M., Cusano, R., Oppo, M., Zoledziewska, M., Pitzalis, M., Deidda, F., Porcu, E., Poddie, F., Kang, H.M., Lyons, R., Tarrier, B., Gresham, J.B., Li, B., Tofanelli, S., Alonso, S., Dei, M., Lai, S., Mulas, A., Whalen, M.B., Uzzau, S., Jones, C., Schlessinger, D., Abecasis, G.R., Sanna, S., Sidore, C., Cucca, F.: Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chrmosome phylogeny. Science **341**, 565–569 (2013)
150. Franklin, A., Edwards, A.W.F., Fairbanks, D.J., Hartl, D.L., Seidenfeld, T.: Ending the Mendel-Fisher Controversy. University of Pittburgh Press, Pittsburgh, PA (2008)
151. Frauenfelder, H., Sligar, S.G., Wolynes, P.G.: The eenergy landscape and motions of proteins. Science **254**, 1598–1603 (1991)
152. Freire, J.G., Field, R.J., Gallas, J.A.C.: Relative abundance and structure of chaotic behavior: The nonpolynomial belousov-zhabotinsky reaction kinetics. J. Chem. Phys. **131**, e044,105 (2009)
153. Fubini, G.: Sugli integrali multipli. Rom. Acc. L. Rend. V **16**, 608–614 (1907). Reprinted in Fubini, G. Opere scelte 2, Cremonese pp. 243-249, 1958
154. Galton, F.: The geometric mean in vital and social statistics. Proc. Roy. Soc. London **29**, 365–367 (1879)
155. Galton, F.: Natural Inheritance, second american edn. Macmillan & Co., London (1889). App. F, pp.241-248
156. Gardiner, C.W.: Handbook of Stochastic Methods, first edn. Springer-Verlag, Berlin (1983)

157. Gardiner, C.W.: Stochastic Methods. A Handbook for the Natural Sciences and Social Sciences, fourth edn. Springer Series in Synergetics. Springer-Verlag, Berlin (2009)
158. Geisler, W., Föllner, H.H.: Three steady state situation in an open chemical reaction system. I. Biophys. Chem. **6**, 107–115 (1977)
159. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: Baysian Data Analysis, second edn. Texts in Statistical Science. Chapman & Hall / CRC, Boca Raton, FL (2004)
160. George, G.: Testing for the independence of three events. Mathematical Gazette **88**, 568 (2004)
161. Georgii, H.: Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik, third edn. Walter de Gruyter GmbH & Co., Berlin (2007). In German. English translation: *Stochastics. Introduction to Probability and Statistics.* Walter de Gruyter GmbH & Co. Berlin (2008).
162. Gibbs, J.W.: Elementary Principles in Statistical Mechanics. Charles Scribner's Sons, New York (1902). Reprinted 1981 by Ox Bow Press, Woodbridge, CT
163. Gibbs, J.W.: The Scientific Papers of J. Willard Gibbs, Vol.I, Thermodynamics. Dover Publications, New York (1961)
164. Gibson, M.A., Bruck, J.: Efficient exact stochastic simulation of chemical systems with many species and many channels. J. Phys. Chem. A **104**, 1876–1889 (2000)
165. Gihman, I.F., Skorohod, A.V.: The Theory of Stochastic Processes. Vol. I, II, and III. Springer-Verlag, Berlin (1975)
166. Gillespie, D.T.: A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comp. Phys. **22**, 403–434 (1976)
167. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. **81**, 2340–2361 (1977)
168. Gillespie, D.T.: Markov Processes: An Introduction for Physical Scientists. Academic Press, San Diego, CA (1992)
169. Gillespie, D.T.: A rigorous derivation of the chemical master equation. Physica A **188**, 404–425 (1992)
170. Gillespie, D.T.: Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral. Phys. Rev. E **54**, 2084–2091 (1996)
171. Gillespie, D.T.: The chemical Langevin equation. J. Chem. Phys. **113**, 297–306 (2000)
172. Gillespie, D.T.: Approximate accelerated stochastic simulation of chemically reacting systems. J. Chem. Phys. **115**(4), 1716–1733 (2001)
173. Gillespie, D.T.: Stochastic simulation of chemical kinetics. Annu. Rev. Phys. Chem. **58**, 35–55 (2007)
174. Gillespie, D.T., Seitaridou, E.: Simple Brownian Diffusion. An Introduction to the Standard Theoretical Models. Oxford University Press, Oxford, UK (2013)
175. Gillies, D.: Varieties of propensity. Brit. J. Phil. Sci. **51**, 807–853 (2000)
176. Goel, N.S., Richter-Dyn, N.: Stochastic Models in Biology. Academic Press, New York (1974)
177. Goychuk, I.: Viscoelastic subdiffusion: Generalized langevin equation approach. Adv. Chem. Phys. **150**, 187–253 (2012)
178. Gradstein, I.S., Ryshik, I.M.: Tables of Series, Products, and Integrals, vol. 1. Verlag Harri Deutsch, Thun, DE (1981). In German and English. Translated from Russian by Ludwig Boll, Berlin
179. Gray, R.M.: Entropy and Information Theory, second edn. Springer, New York (2011)
180. Griffiths, A.J.F., Wessler, S.R., Caroll, J.B., Doebley, J.: An Introduction to Genetic Analysis, tenth edn. W. H. Freeman, New York (2012)

181. Grünbaum, B.: Venn diagrams and independent falilies of sets. Mathematics Magazine **48**, 12–23 (1975)
182. Grünbaum, B.: The construction of Venn diagrams. The College Mathematics Journal **15**, 238–247 (1984)
183. Guckenheimer, J., Holmes, P.: Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, *Applied Mathematical Sciences*, vol. 42. Springer-Verlag, New York (1983)
184. Gunawardena, J.: Chemical reaction network theory for *in-silico* biologists. Tech. rep., Bauer Center for Genomics Research at Harvard University, Cambridge, MA (2003)
185. Györgyi, L., Field, R.J.: A three-variable model of deterministic chaos in the belousov-zhabotinsky reaction. Nature **355**, 808–810 (1992)
186. Hájek, A.: Interpretations of probability. In: E.N. Zalta (ed.) The Stanford Encyclopedia of Philosophy, Winter 2012 edn. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, Stanford Universiy, Stanford, CA. World Wide Web URL: `http://plato.stanford.edu/entries/probability-interpret/` (2013). Retrieved January 23, 2013
187. Hajek, B.: An exploration of random processes for engineers. Lecture Notes ECE 534, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL (2014). Retrieved March 16, 2014 from `www.ifp.illinois.edu/~hajek/Papers/randomprocesses.html`.
188. Hamill, O.P., Marty, A., Neher, E., Sakmann, B., Sigworth, F.J.: Improved patch-clamp techniques for high-resolution current recording from cels and cell-free mambrane patches. Pflügers Archiv. European Journal of Physiology **391**, 85–100 (1981)
189. Hamilton, J.D.: Time Series Analysis. Princeton University Press, Princeton, NJ (1994)
190. Hamilton, W.R.: On a general method in dynamics. Phil. Trans. Roy. Soc. London **II** for 1834, 247–308 (1834)
191. Hamilton, W.R.: Second essay on a general method in dynamics. Phil. Trans. Roy. Soc. London **I** for 1835, 95–144 (1835)
192. Hammer, M.F.: A recent common ancestry for human Y chromosomes. Nature **378**, 376–378 (1995)
193. Hanna, A., Saul, A., Showalter, K.: Detailed studies of propagating frints in the iodate oxidation of arsenous acid. J. Am. Chem. Soc. **104**, 3838–3844 (1982)
194. Hansma, H.G., Kasuya, K., Oroudjev, E.: Atomic force microscopy imaging and pulling of nucleic acids. Curr. Op. Struct. Biol. **14**, 380–385 (2004)
195. Harris, T.E.: Branching Processes. Springer-Verlag, Berlin (1963)
196. Harris, T.E.: The Theory of Branching Processes. Dover Publications, New York (1989)
197. Hartl, D.L., Clark, A.G.: Principles of Population Genetics, third edn. Sinauer Associates, Sunderland, MA (1997)
198. Hartman, P., Wintner, A.: On the law of the iterated logarithm. American Journal of Mathematics **63**, 169–173 (1941)
199. Hatzakis, N.S., Wei, L., Jorgensen, S.K., Kunding, A.H., Bolinger, P.Y., Ehrlich, N., Makarov, I., Skjot, M., Svendsen, A., Hedegård, P., Stamou, D.: Single enzyme studies reveal the existence of discrete states for monomeric enzymes and how they are "selected" upon allosteric regulation. J. Am. Chem. Soc. **134**, 9296–9302 (2012)
200. Haubold, H.J., Mathai, M.A., Saxena, R.K.: Mittag-Leffler functions and their applications. J. Appl. Math. **2011**, e298,628 (2011). Hindawi Publ. Corp.
201. Haussmann, U.G., Pardoux, E.: Time reversal of diffusions. Annals Probability **14**, 1188–1205 (1986)

202. Hawkins, D., Ulam, S.: Theory of multiplicative processes I. Tech. Rep. LADC-265, Los Alamos Scientific Laboratory (1944)
203. Hazeltine, E.L., Rawlings, J.B.: Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. J. Chem. Phys. **117**, 6959–6969 (2002)
204. Heathcote, C.R., Moyal, J.E.: The random walk (in continuous time) and its application to the theory of queues. Biometrika **46**, 400–411 (1959)
205. Heinrich, R., Sonntag, I.: Analysis of the selection equation for a multivariable population model. Deterministic anad stochastic solutios and discussion of the approach for populations of self-reproducing biochemical networks. J. theor. Biol. **93**, 325–361 (1981)
206. Heyde, C.C., Seneta, E.: Studies in the history of probability and statistics. xxxi. the simple branching porcess, a turning point test and a fundmanetal inequality: A historical note on I. J. Bienaymé. Biometrika **59**, 680–683 (1972)
207. Higham, D.J.: Modeling and somulationg chemical reactions. SIAM Review **50**, 347–368 (2008)
208. Hinshelwood, C.N.: On the theory of unimolecular reactions. Proc. Roy. Soc. LOndon A **113**, 230–233 (1926)
209. Hirsch, M.W., Smale, S.: Differential Equations, Dynamical Systems, and an Introduction to Chaos, second edn. Elsevier, Amsterdam (2004)
210. Hirschfeld, T.: Optical microscopic observation of small molecules. Applied Optics **15**, 2965–2966 (1976)
211. Hocking, R.L., Schwertman, N.C.: An extension of the birthday problem to exactly k matches. The College Mathematics Journal **17**, 315–321 (1986)
212. Hogg, R.V., McKean, J.W., Craig, A.T.: Introduction to Mathematical Statistics, seventh edn. Pearson Education, Upper Saddle River, NJ (2012)
213. Hogg, R.V., Tanis, E.A.: Probability and Statistical Inference, eighth edn. Pearson – Prentice Hall, Upper Saddle River, NJ (2010)
214. Holder, M., Lewis, P.O.: Phylogeny estimation: Traditional and Bayesian approaches. Nature Reviews Genetics **4**, 275–284 (2003)
215. Holdren, J.P., Lander, E., Varmus, H.: Designing a Digital Future: Federally Funded Research and Development in Networking and Information Technology. President's Council of Advisors on Science and Technology, Washington, DC (2010)
216. Holsinger, K.E.: Lecture Notes in Population Genetics. University of Connecticut, Dept. of Ecology and Evolutionary Biology, Storrs, CT (2012). Licensed under the Creative Commons Attribution-ShareAlike License: `http://creativecommons.org/licenses/by-sa/3.0/`
217. Horn, F.: Necessary and sufficient conditions for complex balancing in chemical kinetics. Archive for Rational Mechanics and Analysis **49**, 172–186 (1972)
218. Horn, F., Jackson, R.: General mass action kinetics. Archive for Rational Mechanics and Analysis **47**, 81–116 (1972)
219. Houchmandzadeh, B., Vallade, M.: An alternative to the diffusion equation in population genetics. Phys. Rev. E **83**, e051,913 (2010)
220. Houston, P.L.: Chemical Kinetics and Reaction Dynamics. The McGraw-Hill Companies, New York (2001)
221. Hu, Y., Li, T.: Highly accurate tau-leaping methiods with random corrections. J. Chem. Phys. **130**, e124,109 (2009)
222. Hua, Y., Sarkar, T.K.: Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. IEEE Trans. Acoustics, Speech, and Signal Processing **38**, 814–824 (1990)
223. Humphries, N.E., Queiroz, N., Dyer, J.R.M., Pade, N.G., Musyl, M.K., Schaefer, K.M., Fuller, D.W., Brunnschweiler, J.M., Doyle, T.K., Houghton, J.D.R., Hays, G.C., Jones, C.S., Noble, L.R., Wearmouth, V.J., Southall, E.J., Sims, D.W.: Environmental context explains Lévy and Brwonian movement patterns of marine predators. Nature **465**, 1066–1069 (2010)

224. Inagaki, H.: Selection under random mutations in stochastic Eigen model. Bull. Math. Biol. **44**, 17–28 (1982)
225. Ishida, K.: Stochastic model for bimolecular reaction. J. Chem. Phys. **41**, 2472–2478 (1964)
226. Itō, K.: Stochastic integral. Proc. Imp. Acad. Tokyo **20**, 519–524 (1944)
227. Itō, K.: On stochastic differential equations. Mem. Amer. Math. Soc. **4**, 1–51 (1951)
228. Jachimowski, C.J., McQuarrie, D.A., Russell, M.E.: A stochastic approach to enzyme-substrate reactions. Biochemistry **3**, 1732–1736 (1964)
229. Jackson, E.A.: Perspectives of Nonlinear Dynamics, vol. 1. Cambridge University Press, Cambridge, UK (1989)
230. Jackson, E.A.: Perspectives of Nonlinear Dynamics, vol. 2. Cambridge University Press, Cambridge, UK (1989)
231. Jacobs, K.: Stochastic processes for Physicists. Understanding Noisy Systems. Cambridge University Press, Cambridge, UK (2010)
232. Jaynes, E.T.: Information theory and statistical mechanics. Phys. Rev. **106**, 620–630 (1957)
233. Jaynes, E.T.: Information theory and statistical mechanics. II. Phys. Rev. **108**, 171–190 (1957)
234. Jaynes, E.T.: Probability Theory. The Logic of Science. Cambridge University Press, Cambridge, UK (2003)
235. Johnson, N.L., Kotz, S., Balakrishnan, N.: Continuous Univariate Distributions, *Probability and Mathematical Statistics. Applied Probability and Statistics*, vol. 1, second edn. John Wiley & Sons, New York (1994)
236. Johnson, N.L., Kotz, S., Balakrishnan, N.: Continuous Univariate Distributions, *Probability and Mathematical Statistics. Applied Probability and Statistics*, vol. 2, second edn. John Wiley & Sons, New York (1995)
237. Jones, B.L., Enns, R.H., Rangnekar, S.S.: On the theory of selection of coupled macromolecular systems. Bull. Math. Biol. **38**, 15–28 (1976)
238. Jones, B.L., Leung, H.K.: Stochastic analysis of a non-linear model for selection of biological macromolecules. Bull. Math. Biol. **43**, 665–680 (1981)
239. Joyce, G.F.: Forty years of *in vitro* evolution. Angew. Chem. Internat. Ed. **46**, 6420–6436 (2007)
240. Kassel, L.S.: Studies in homogeneous gas reactions I. J. Phys. Chem. **32**, 225–242 (1928)
241. Kendall, D.G.: An artificial realization of a simple "birth-and-death" process. J. Roy. Statist. Soc. B **12**, 116–119 (1950)
242. Kendall, D.G.: Branching processes since 1873. J. of the London Mathematical Society **41**, 386–406 (1966)
243. Kendall, D.G.: The genalogy of genealogy: Branching processes before (an after) 1873. Bull. of the London Mathematical Society **7**, 225–253 (1975)
244. Kenney, J.F., Keeping, E.S.: Mathematics of Statistics, second edn. Van Nostrand, Princeton, NJ (1951)
245. Kenney, J.F., Keeping, E.S.: The k-Statistics. In Mathematics of Statistics. Part I, § 7.9, third edn. Van Nostrand, Princeton, NJ (1962)
246. Kesten, H., Stigum, B.P.: A limit theorem for multidimensional Galton-Watson processes. Annal. Math. Statistics **37**, 1211–1223 (1966)
247. Keynes, J.M.: A Treatise on Probability. MacMillan & Co., London (1921)
248. Khinchin, A.Y.: Über einen Satz der Wahrscheinlichkeitsrechnung. Fundamenta Mathematica **6**, 9–20 (1924). In German
249. Kim, S.K.: Mean first passage time for a random walker and its application to chemical knietics. J. Chem. Phys. **28**, 1057–1067 (1958)
250. Kimura, M.: Solution of a process of random genetic drift with a continuous model. Proc. Natl. Acad. Sci. USA **41**, 144–150 (1955)

251. Kimura, M.: Diffusion models in population genetics. J. Appl. Prob. **1**, 177–232 (1964)
252. Kimura, M.: The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, UK (1983)
253. Kingman, J.F.C.: Mathematics of Genetic Diversity. Society for Industrial and Applied Mathematics, Washigton, DC (1980)
254. Kingman, J.F.C.: The genealogy of large populations. J. Appl. Prob. **19**(Essays in Statistical Science), 27–43 (1982)
255. Kingman, J.F.C.: Origins of the coalescent: 1974 – 1982. Genetics **156**, 1461–1463 (2000)
256. Knuth, D.E.: Two notes on notation. Am. Math. Monthly **99**, 403–422 (1992)
257. Kolmogorov, A.N.: Über das Gesetz es interierten Logarithmus. Mathematische Annalen **101**, 126–135 (1929). In German
258. Kolmogorov, A.N.: Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. Mathematische Annalen **104**, 415–458 (1931)
259. Kolmogorov, A.N.: Grundbegriffe der Wahrscheinlichkeitsrechnung. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer-Verlag, Berlin (1933). English translation: *Foundations of Probability*. Chelsea Publ. Co. New York, 1950
260. Kolmogorov, A.N., Dmitriev, N.A.: "Zur Lösung einer biologischen Aufgabe". Isvestiya Nauchno-Issledovatel'skogo Instituta Matematiki i Mekhaniki pri Tomskom Gosudarstvennom Universitete **2**, 1–12 (1938)
261. Kolmogorov, A.N., Dmitriev, N.A.: Branching stochastic processes. Doklady Akad. Nauk U.S.S.R. **56**, 5–8 (1947)
262. Kou, S.C., Cherayil, B.J., Min, W., English, B.P., Xie, X.S.: Single-molecule Michaelis-Menten equations. J. Phys. Chem. B **109**, 19,068–19,081 (2005)
263. Kowalski, C.J.: Non-normal bivariate distributions with normal marginals. The American Statistician **27**, 103–106 (1973)
264. Krichevsky, O., Bonnet, G.: Fluorescence correlation spectroscopy: The technique and its applications. Rep. Prog. Phys. **65**, 251–297 (2002)
265. Kubo, R.: The fluctuation-dissipation theorem. Rep. Prog. Phys. **29**, 255–284 (1966)
266. Kügler, P., Gaubitzer, E., Müller, S.: Perameter identification for chemical reaction systems using sparsity enforcing regularization A case study for the chlorite–iodide reaction. J. Phys. Chem. A **113**, 2775–2785 (2009)
267. Kulzer, F., Orrit, M.: Single-molecule optics. Annu. Rev. Phys. Chem. **55**, 585–611 (2004)
268. Kumaresan, R., Tufts, D.W.: Estimating the parameters of exponentially damped sinusoids and pole-zero modeiling in noise. IEEE Trans. Acoustics, Speech, and Signal Processing **30**, 833–840 (1982)
269. Laidler, K.J.: Chemical Kinetics, third edn. Addison Wesley, Boston, MA (1987)
270. Laidler, K.J., King, M.C.: The development of transition-state theory. J. Phys. Chem. **87**, 2657–2664 (1983)
271. Langevin, P.: Sur la théorie du mouvement Brownien. Comptes Rendues hebdomadaires des Séances de L'Academié des Sceinces **146**, 530–533 (1908)
272. Laplace, P.S.: Essai philosophique les probabililtés. Courcies Imprimeur, Paris (1814). English edition: A Philosophical Eaasay on Probabilties. Dover Publications, New York, 1951
273. Laurenzi, I.J.: An analytical solution of the stochastic master equation for reversible bimolecular reaction kinetics. J. Chem. Phys. **113**, 3315–3322 (2000)
274. Lauritzen, S.L.: Time series analysis in 1880: A discussion of contributions made by t. n, thiele. International Statistical Review **49**, 319–331 (1981)
275. Lee, P.M.: Bayesian Statistics, third edn. Hodder Arnold, London (2004)

276. Leemis, L.: Poisson to normal. College of William & Mary, Department of Mathematics, Williamsburg, VA (2012). URL: www.math.wm.edu/~leemis/chart/UDR/PDFs/PoissonNormal.pdf
277. Lefever, R., Nicolis, G., Borckmans, P.: The Brusselator: It does oscillate all the same. J. Chem. Soc., Faraday Trans. 1 pp. 1013–1023 (1988)
278. Legendre, A.M.: Nouvelles méthodes pour la détermination des orbites des comètes. F. Didot, Paris (1805). In French
279. Lerch, H.P., Rigler, R., Mikhailov, A.S.: Functional conformational motions in the turnover cycle of cholesterol oxidase. Proc. Natl. Acad. Sci. USA **102**, 10,807–10,812 (2005)
280. Lévy, P.: Calcul de probabilités. Geuthier-Villars, Paris (1925). In French
281. Li, H., Cao, Y., Petzold, L.R., Gillespie, D.T.: Algortihms and software for stochastic simulation of biochemical reacting systems. Biotechnol Prog. **24**, 56–61 (2008)
282. Li, P.T.X., Bustamante, C., Tinoco, Jr., I.: Real-time control of the energy landscape by force directs the folding of RNA molecules. Proc. Natl. Acad. Sci. USA **104**, 7039–7044 (2007)
283. Li, T.: Analysis of explicit tau-leaping schemes for simulating chemically reacting systems. Multiscale Model. Simul. **6**, 417–436 (2007)
284. Li, T., Kheifets, S., Medellin, D., Raizen, M.G.: Measurement of the instantaneous velocity of a Brownian particle. Science **328**, 1673–1675 (2010)
285. Liao, D., Galajda, P., Riehn, R., Ilic, R., Puchalla, J.L., Yu, H.G., Craighead, H.G., Austin, R.H.: Single molecule correlation spectroscopy ib continuous flow mixers with zero-mode waveguides. Optics Express **16**, 10,077–10,090 (2008)
286. Limpert, E., Stahel, W.A., Abbt, M.: Log-normal distributions across the sciences: Keys and clues. BioScience **51**, 341–352 (2001)
287. Lin, H., Truhlar, D.G.: QM/MM: What have we learned, where are we, and where do we go from here? Theor. Chem. Acc. **117**, 185–199 (2007)
288. Lin, S.H., Lau, K.H., Richardson, W., Volk, L., Eyring, H.: Stochastic model of unimolecular reactions and the RRKM theory. Proc. Nat. Acad. Sci. USA **69**, 2778–2782 (1972)
289. Lindeberg, J.W.: Über das Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. Annales Academiae Scientiarum Fennicae **16**, 1–23 (1920). In German.
290. Lindeberg, J.W.: Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. Mathematische Zeitschrift **15**, 211–225 (1922). In German
291. Lindemann, F.A.: Discussion on the radiation theory on chemical action. Trans. Farad. Soc. **17**, 598–606 (1922)
292. Liouville, J.: Note sur la théorie de la variation des constantes arbitraires. Journal de Mathématiques pure et appliquées **3**, 342–349 (1838). In French.
293. Liouville, J.: Mémoire sur l'intégration des équations différentielles du mouvement quelconque de points matériels. Journal de Mathématiques pure et appliquées **14**, 257–299 (1849). In French.
294. Lorenz, E.N.: Deterministic nonperiodic flow. J. Atmospheric Sciences **20**, 130–141 (1963)
295. Lu, H.P., Xun, L., Xie, X.S.: Single-molecule enzyme dynamics. Science **282**, 1877–1882 (1998)
296. Lu, J., Engl, H.W., Rainer Machné, Schuster, P.: Inverse bifurcation analysis of a model for the mammalian G1/S regulatory module. Lecture Notes in Computer Science **4414**, 168Ű184 (2007)
297. Lu, J., Engl, H.W., Schuster, P.: Inverse bifurcation analysis: Application to simple gene systems. ABM – Algorithms for Molecular Biology **1**, e11 (2006)
298. Lyapunov, A.M.: Sur une proposition de la théorie des probabilités. Bull. Acad. Imp. Sci. St. Pétersbourg **13**, 359–386 (1900)

299. Lyapunov, A.M.: Nouvelle forme du théorème sur la limite des probabilités. Mem. Acad. Imp. Sci. St. Pétersbourg, Classe Phys. Math. **12**, 1–24 (1901)
300. Magde, D., Elson, E., Webb, W.W.: Thermodynamic fluctuations in a reating system – Measurement by fluorescence correlation spectroscopy. Phys. Rev. Letters **29**, 705–708 (1972)
301. Mahnke, R., Kaupužs, J., Lubashevsky, I.: Physics of Stochastic Processes. How Randomness Acts in Time. Wiley-VCh Verlag, Weinheim (Bergstraße), DE (2009)
302. Mallows, C.: Anothre comment on O'Cinneide. The American Statistician **45**, 257 (1991)
303. Mandelbrot, B.B.: The Fractal Geometry of Nature, updated edn. W. H. Freeman Company, New York (1983)
304. Mansuy, R.: The origins of the word "martingale". Electronic Journal for History of Probability and Statistics **5**(1), 1–10 (2009). Translated by Ronald Sverdlove from the French *Histoire des martigales*. Mathématiques & Sciences Humaines 43(169), 105-113, 2005
305. Marcus, R.A.: Unimolecular dissociations and free radical recombination reactions. J. Chem. Phys. **20**, 359–364 (1952)
306. Marcus, R.A.: Vibrational nonadiabaticity and tunneling effects in thranition state theory. J. Chem. Phys. **83**, 204–207 (1979)
307. Marcus, R.A.: Unimolecular reactions, rates and quantum state distributions of products. Phil. Trans. Roy. Soc. London A **332**, 283–296 (1990)
308. Marcus, R.A., Rice, O.K.: The kinetics of the recombination of methyl radical and iodine atoms. J. Phys. Colloid Chem. **55**, 894–908 (1951)
309. Maruyama, T.: Stochastic Problems in Population Genetics. Springer-Verlag, Berlin (1977)
310. Marx, D., Jürg Hutter: *Ab initio* Molecular Dynamics. Basic Theory and Advanced Methods. Cambridge University Press, Cambridge, UK (2009)
311. Mathai, A.M., Saxena, R.K., Haubold, H.J.: A certain class of Laplace transforms with applications to reaction and reaction-diffusion equations. Astrophys. Space Sci. **305**, 283–288 (2006)
312. McAlister, D.: The law of the geometric mean. Proc. Roy. Soc. London **29**, 367–376 (1879)
313. McCaskill, J.S.: A stochastic theory of macromolecular evolution. Biol. Cybern. **50**, 63–73 (1984)
314. McKean, Jr., H.P.: Stochastic Integrals. John Wiley & Sons, New York (1969)
315. McQuarrie, D.A.: Kinetics of small systems. I. J. Chem. Phys. **38**, 433–436 (1962)
316. McQuarrie, D.A.: Stochastic approach to chemical kinetics. J. Appl. Prob. **4**, 413–478 (1967)
317. McQuarrie, D.A.: Mathematical Methods for Scientists and Engineers. University Science Books, Sausalito, CA (2003)
318. McQuarrie, D.A., Jachimowski, C.J., Russell, M.E.: Kinetics of small systems. II. J. Chem. Phys. **40**, 2914–2921 (1964)
319. Medina, M.Ángel., Schwille, P.: Fluorescence correlation spectroscopy for the detection and study of single molecules in biology. BioEssays **24**, 758–764 (2002)
320. Medvegyev, P.: Stochastic Integration Theory. Oxford University Press, New York (2007)
321. Meinhardt, H.: Models of Biological Pattern Formation. Academic Press, London (1982)
322. Meintrup, D., Schäffler, S.: Stochastik. Theorie und Anwendungen. Springer-Verlag, Berlin (2005). In German
323. Melnick, E.L., Tenenbein, A.: Misspecifications of the normal distribution. The American Statistician **36**, 372–373 (1982)

324. Mendel, G.: Versuche über Pflanzen-Hybriden. Verhandlungen des naturforschenden Vereins in Brünn **IV**, 3–47 (1866). In German
325. Meredith, M.: Born in Africa: The Quest for the Origins of Human Life. Public Affairs, New York (2011)
326. Merkle, M.: Jensen's inequality for medians. Statistics & Probability Letters **71**, 277–281 (2005)
327. Messiah, A.: Quantum Mechanics, vol. II. North-Holland Publishing Company, Amsterdam, NL (1970). Translated from the French by J. Potter
328. Metzler, R., Klafter, J.: The random walk's guide to anomalous diffusion: A fractional dynamics approach. Physics Reports **339**, 1–77 (2000)
329. Michaelis, L., Menten, M.L.: The kinetics of the inversion effect. Biochemische Zeitschrift **49**, 333–369 (1913)
330. Miller, R.W.: Propensity: Popper or Peirce? Brit. J. Phil. Sci. **26**, 123–132 (1975)
331. Mittag-Leffler, M.G.: Sur la nouvelle fonction $E_\alpha(x)$. C. R. Acad. Sci. Paris, Ser.II, **137**, 554–558 (1903)
332. Moerner, W.E., Kador, L.: Optical detection and spectroscopy of single molecules in a solid. Phys. Rev. Lett. **62**, 2535–2538 (1989)
333. Monod, J., Wyman, J., Changeaux, J.P.: On the natur of allosteric transitions: A plausible model. J. Mol. Biol. **12**, 88–118 (1965)
334. Montroll, E.W.: Stochastic processes and chemical kinetics. In: W.M. Muller (ed.) Energetics in Metallurgical Phenomenon, vol. 3, pp. 123–187. Gordon & Breach, New York (1967)
335. Montroll, E.W., Shuler, K.E.: Studies in nonequilibrium rate processes: I. The relaxation of a system of harmonic oscillators. J. Chem. Phys. **26**, 454–464 (1956)
336. Montroll, E.W., Shuler, K.E.: The application of the theory of stochastic processes to chemical kinetics. Adv. Chem. Phys. **1**, 361–399 (1958)
337. Montroll, E.W., Weiss, G.H.: Random walks on lattices. II. J. Math. Phys. **6**, 167–181 (1965)
338. Moore, G.E.: Cramming more components onto intergrated circuits. Electronics **38**(8), 4–7 (1965)
339. Moran, P.A.P.: Random processes in genetics. Proc. Camb. Phil. Soc. **54**, 60–71 (1958)
340. Moran, P.A.P.: The Statistical Processes of Evolutionary Theroy. Clarendon Press, Oxford, UK (1962)
341. Motulsky, H.J., Christopoulos, A.: Fitting Models to Biological Data Using Linear and Nonlinear Regression. A Practical Guide to Curve Fitting. GraphPad Software Inc., San Diego, CA (2003)
342. Mount, D.W.: Bioinformatics. Sequence and Genome Analysis, second edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (2004)
343. Moyal, J.E.: Stochastic processes and statistical physics. J. Roy. Statist. Soc. Ḃ **11**, 150–210 (1949)
344. Müller, S., Regensburger, G.: Generalized mass action systems: Complex balanding equilibria and sign vectors of the stoichiometric and kinetic-order subspaces. SIAM J. Appl. Math. **72**, 1926–1947 (2012)
345. Neher, E., Sakmann, B.: Single-cheannel currents recorded from membrane of denervated frog muscle fibres. Nature **260**, 799–802 (1976)
346. Nicolis, G., Prigogine, I.: Self-Organization in Nonequilibrium Systems. John Wiley & Sons, New York (1977)
347. Nolan, J.P.: Stable Distributions: Models for Heavy-Tailed Data. Birkhäuser, Boston, MA (2013). Unfinished manuscript. Online at `academic2.american.edu/∼jpnolan`.
348. Novitski, C.E.: On Fisher's criticism of Mendel's results with the garden pea. Genetics **166**, 1133–1136 (2004)

349. Novitski, C.E.: Revision of Fisher's analysis of Mendel's garden pea experiments. Genetics **166**, 1139–1140 (2004)
350. Noyes, R.M., Field, R.J., Körös, E.: Oscillations in chemical systems. I. Detailed mechanism in a system showing temporal oscillations. J. Am. Chem. Soc. **94**, 1394–1395 (1972)
351. Nyman, J.E.: Another generalization of the birthday problem. Mathematics Magazine **48**, 46–47 (1975)
352. Øksendal, B.K.: Stochastic Differential Equations. An Introduction with Applications, sixth edn. Springer-Verlag, Berlin (2003)
353. Olbregts, J.: Termolecular reaction of nitrogen monoxide and oxygen. A still unsolved problem. Internat. J. Chem. Kinetics **17**, 835–848 (1985)
354. Onuchic, J.N., Luthey-Schulten, Z., Wolynes, P.G.: Theory of protein folding: The energy landscape perspective. Annu. Rev. Phys. Chem. **48**, 545–600 (1997)
355. Orrit, M., Bernard, J.: Single pentacene molecules detected by fluorescence exitation in a p-terphenyl crystal. Phys. Rev. Lett. **65**, 2716–2719 (1990)
356. Oster, G.F., Perelson, A.S.: Chemical reaction dynamics. Part I: Geometrical structure. Arch. Rational Mech. Anal. **55**, 230–274 (1974)
357. Park, S.Y., Bera, A.K.: Maximum entropy autoregressive conditional heteroskedasticy model. J. Econometrics **150**, 219–230 (2009)
358. Paschotta, R.: Field Guide to Laser Puls Generation. SPIE Press, Bellingham, WA (2008)
359. Patrick, R., Golden, D.M.: Third-order rate constants of atmospheric importance. Internat. J. Chem. Kinetics **15**, 1189–1227 (1983)
360. Pearson, E.S., Wishart, J.: "Student's" Collected Papers. Cambridge University Press, Cambridge, UK (1942). Cambridge University Press for the Biometrika Trustees
361. Pearson, J.A.: Advanced Statistical Physics. University of Manchester, Manchester, UK (2009). URL: `http://www.joffline.com/`
362. Pearson, K.: Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. Phil. Trans. Roy. Soc. London A **186**, 343–414 (1895)
363. Pearson, K.: On the criterion that a given system of deviations form the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine Series 5 **50**(302), 157–175 (1900)
364. Pearson, K.: The problem of the random walk. Nature **72**, 294 (1905)
365. Peirce, C.S.: Vol.7: Science and philosophy and Vol.8: Reviews, correspondence, and bibliography. In: A.W. Burks (ed.) The Collected Papers of Charles Sanders Peirce, vol. 7-8. Belknap Press of Harvard University Press, Cambridge, MA (1958)
366. Peterman, E.J.G., Sosa, H., Moerner, W.E.: Single-molecule flourescence spectrocopy and microscopy of biomolecular motors. Annu. Rev. Phys. Chem. **55**, 79–96 (2004)
367. Philibert, J.: One and a half century of diffusion: Fick, Einstein, before and beyond. Diffusion Fundamentals **4**, 6.1–6.19 (2006)
368. Phillipson, P.E., Schuster, P.: Modeling by Nonlinear Differential Equations. Dissipative and Conservative Processes, *World Scientific Series on Nonlinear Science A*, vol. 69. World Scientific, Singapore (2009)
369. Plass, W.R., Cooks, R.G.: A model for energy transfer in inelasitc molecular collisions applicable at steady state and non-steady state and for an arbitrary distribution of collision energies. J. Am. Soc. Mass Spectrom. **14**, 1348–1359 (2003)
370. Pollard, H.: The representatioin of e^{-x^λ} as a Laplace intgeral. Bull. Am. Math. Soc. **52**, 908–910 (1946)

371. Popper, K.: The propensity interpretation of the calculus of probability and of the quantum theory. In: S. Körner, M.H.L. Price (eds.) Observation and Interpretation in the Philosophy of Physics: Proceedings of the Ninth Symposium of the Colston Research Society. Butterworth Scientific Publications, London (1957)
372. Popper, K.: The propensity theory of probability. Brit. J. Phil. Sci. **10**, 25–62 (1960)
373. Poznik, G.D., Henn, B.M., Yee, M.C., Sliwerska, E., Lin, A.A., Snyder, M., Quintana-Murci, L., Kidd, J.M., Underhill, P.A., Bustamante, C.D.: Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. Science **341**, 562–565 (2013)
374. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T.: Numerical Recipes. The Art of Scientific Computing. Cambridge University Press, Cambridge, UK (1986)
375. Price, R.: LII. an essay towards soliving a problem in the doctrine of chances. By the late Ref. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M.A. and F.R.S. Phil. Trans. Roy. Soc. London **53**, 370–418 (1763)
376. Protter, P.E.: Stochastic Intergration and Differential Equations, *Applications of Mathematics*, vol. 21, second edn. Springer-Verlag, Berlin (2004)
377. Provencher, S.W., Dovi, V.G.: Direct analysis of continuous relaxation spectra. J. Biophys. Biochem. Methods **1**, 313–318 (1979)
378. Qian, H., Elson, E.L.: Single-molecule enzymology: Stochastic Michaelis-Menten kinetics. Biophys. Chem. **101-102**, 565–576 (2002)
379. Rathinam, M., Petzold, L.R., Cao, Y., Gillespie, D.T.: Stiffness in stochastic chemically reacting systems: The implicit τ-leaping method. J. Chem. Phys. **119**, 12,784–12,794 (2003)
380. Rice, O.K., Ramsperger, H.C.: Theories of unimolecular gas reactions at low pressures. J. Am. Chem. Soc. **49**, 1617–1629 (1927)
381. Rigler, R., Mets, U., Widengren, J., Kask, P.: Fluorescence correlation spectroscopy with high count rate and low-background-analysis of translational diffusion. Eur. Biophys. J. **22**, 169–175 (1993)
382. Riley, K.F., Hobson, M.P., Bence, S.J.: Mathematical Methods for Physics and Engineering, second edn. Cambridge University Press, Cambridge, UK (2002)
383. Risken, H.: TheFokker-Planck Equation. Methods of Solution and Applications, 2nd edn. Springer-Verlag, Berlin (1989)
384. Robinett, R.W.: Quantum Mechanics. Classical Results, Modern Systems, and Visualized Examples. Oxford University Press, New York (1997)
385. Roebuck, J.R.: The rate of the reaction between arsenious acid and iodine in acid solution, the rate of the reverse reaction, and the equilibrium between them. J. Phys. Chem. **6**, 365–398 (1901)
386. Rotman, B.: Measurement of activity of single molecules of β-d-galactosidase. Proc. Natl. Acad. Sci. USA **47**, 1981–1991 (1961)
387. Sagués, F., Epstein, I.R.: Nonlinear chemical dynamics. J. Chem. Soc., Dalton Trans. **2003**, 1201–1217 (2003)
388. Sanft, K.R., Wu, S., Roh, M., Fu, J., Lim, R.K., Petzold, L.R.: StochKit2: Software for discrete stochastic simulation of biochemical systems with events. Bioinformatics **27**, 2457–2458 (2011)
389. Scher, H., Shlesinger, M.F., Bendler, J.T.: Time scale invariance in transport and relaxation. Physics Today **44**(1), 26–34 (1991)
390. Schilling, M.F., Watkins, A.E., Watkins, W.: Is human height bimodal? The American Statistician **56**, 223–229 (2002)
391. Schlögl, F.: Chemical reaction models for non-equilibrium phase transitions. Z. Physik **253**, 147–161 (1972)
392. Schoutens, W.: Lévy Processes in Finance. Wiley Series in Probability and Statistics. John Wiley & Sons, Chichester, UK (2003)

393. Schubert, M., Weber, G.: Quantentheorie. Grundlagen und Anwendungen. Spektrum Akademischer Verlag, Heidelberg, DE (1993). In German
394. Schuster, P.: Are computer scientists the sutlers of modern biology? Bioinformatics is indispesible for progress in molecular life sciences but does not get credit for its contributions. Complexity **19**(4), 10–14 (2014)
395. Schuster, P., Sigmund, K.: Random selection - A simple model based on linear birth and death processes. Bull. Math. Biol. **46**, 11–17 (1984)
396. Schwabl, F.: Quantum Mechanics, 4th edn. Springer-Verlag, Berlin (2007)
397. Schwarz, G.: Kinetic analysis by chemical relaxation methods. Rev. Mod. Phys. **40**, 206–218 (1968)
398. Sehl, M., Alekseyenko, A.V., Lange, K.L.: Accurate stochastic simulation via the step anticipation τ-leaping (SAL) algorithm. J. Comp.,Biol. **16**, 1195–1208 (2009)
399. Selmeczi, D., Tolić-Nørrelykke, S., Schäffer, E., Hagedorn, P.H., Mosler, S., Berg-Sørensen, K., Larsen, N.B., Flyvbjerg, H.: Brownian motion after Einstein: Some new applications and new experiments. Lect. Notes Phys. **711**, 181–199 (2007)
400. Seneta, E.: Non-negative Matrices and Markov Chains, second edn. Springer-Verlag, New York (1981)
401. Seneta, E.: The central limit problem and lienear least squares in pre-revolutionary Russia: The background. Mathematical Scientist **9**, 37–77 (1984)
402. Senn, H.M., Thiel, W.: QM/MM Methods for biological systems. Top. Curr. Chem. **268**, 173–290 (2007)
403. Senn, H.M., Thiel, W.: QM/MM Methods for biomolecular systems. Angew. Chem. Int. Ed. **48**, 1198–1229 (2009)
404. Seydel, R.: Practical Bifurcation and Stability Analysis. From Equilibrium to Chaos, *Interdisciplinary Applied Mathematics*, vol. 5, second edn. Springer-Verlag, New York (1994)
405. Shannon, C.E.: A mathematical theory of communication. Bell System Technical Journal **27**, 379–423 (1948)
406. Shannon, C.E., Weaver, W.: The Mathematical Theory of Communication. Univ. of Illinois Press, Urbana, IL (1949)
407. Shapiro, B.E., Levchenko, A., World, E.M.M.B.J., Mjolsness, E.D.: Cellerator: Extending a computer algebra system to include biochemical arrows for signal transduction simulations. Bioinformatics **19**, 677–678 (2003)
408. Sharpe, M.J.: Transformations of diffusion by time reversal. The Annals of Probability **8**, 1157–1162 (1980)
409. Shuler, K.E.: Studies in nonequilibrium rate processes: II. The relaxation of vibrational nonequilibrium distributions in chemical reactions and shock waves. J. Phys. Chem. **61**, 849–856 (1957)
410. Shuler, K.E., Weiss, G.H., Anderson, K.: Studies in nonequilibrium rate processes. V. The relaxation of moments derived from a master equation. J. Math. Phys. **3**, 550–556 (1962)
411. Sotiropoulos, V., Kaznessis, Y.N.: Analytical derivation of moment equations in stochastic chemical kinetics. Chem. Eng. Sci. **66**, 268–277 (2011)
412. Stauffer, P.H.: Flux flummoxed: A proposal for consistent usage. Ground Water **44**, 125–128 (2006)
413. Steffensen, J.F.: "deux problème du calcul des probabilités". Ann. Inst. Henri Poincaré **3**, 319–344 (1933)
414. Stepanow, S., Schütz, G.M.: The distribution function os a semiflexible polymer and random walks with constraints. Europhys. Letters **60**, 546–551 (2002)
415. Stevens, J.W.: What is Bayesian Statistics? What is ... ? Hayward Medical Communications, a division of Hayward Group Ltd., London (2009)

416. Strang, G.: Linear Algebra and its Applications, third edn. Harcourt Brace Jovanovich (1988)
417. Stratonovich, R.L.: Introduction to the Theory of Random Noise. Gordon and Breach, New York (1963)
418. Strogatz, S.H.: Nonlinear Dynamics and Chaos. With Applications to Physics, Biology, Chemistry, and Engineering. Westview Press at Perseus Books, Cambridge. MA (1994)
419. Stuart, A., Ord, J.K.: Kendall's Advanced Theory of Statistics. Volume 1: Distribution Theory, fifth edn. Charles Griffin & Co., London (1987)
420. Stuart, A., Ord, J.K.: Kendall's Advanced Theory of Statistics. Volume 2: Classical Inference and Relationship, fifth edn. Edward Arnold, London (1991)
421. Student: The probable error of a mean. Biometrika **6**, 1–25 (1908)
422. Swamee, P.K.: Near lognormal distribution. J. Hydrological Engineering **7**, 441–444 (2007)
423. Swetina, J., Schuster, P.: Self-replication with errors - A model for polynucleotide replication. Biophys. Chem. **16**, 329–345 (1982)
424. Tang, H., Siegmund, D.O., Shen, P., Oefner, P.J., Feldman, M.W.: Frequentist estimation of coalescence times from nucleotide sequence data using a tree-based partition. Genetics **161**, 448–459 (2002)
425. Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation. Society for Industrial and Applied Mathematics, Philadelphia, PA (2005)
426. Tavaré, S.: Line-of-descent and genealogical processes, and their application in population genetics models. Theor. Popul. Biol. **26**, 119–164 (1984)
427. Taylor, H.M., Karlin, S.: An Introduction to Stochastic Modeling, third edn. Academic press, San Diego, CA (1998)
428. Thiele, T.N.: Om Anvendelse af midste Kvadraters Methode i nogle Tilfælde, hvor en Komplikation af visse Slags uensartede tilfædige Feijlkilder giver Feijlene en 'systenatisk' Karakter. Vidensk. Selsk. Skr. 5. rk., naturvid. og mat. Afd. **12**, 381–408 (1880). In Danish.
429. Thompson, C.J., McBride, J.L.: On Eigen's theory of the self-organization of matter and the evolution of biological macromolecules. Math. Biosci. **21**, 127–142 (1974)
430. Tolman, R.C.: The Principle of Statistical Mechanics. Oxford University Press, Oxford, UK (1938)
431. Tsukahara, H., Ishida, T., Mayumi, M.: Gas-phase oxidation of nitric oxide: Chemical kinetics and rate constant. Nitric Oxide: Biology and Chemistry **3**, 191–198 (1999)
432. Turing, A.M.: The chemical basis of morphogenesis. Phil. Trans. Roy. Soc. London B **237**(641), 37–72 (1952)
433. Uhlenbeck, G.E., Ornstein, L.S.: On the theory of the Brownian motion. Phys. Rev. **36**, 823–841 (1930)
434. Ullah, M., Wolkenhauer, O.: Family tree of Markov models in systems biology. IET Systems Biology **1**, 247–254 (2007)
435. Ullah, M., Wolkenhauer, O.: Stochastic Approaches for Systems Biology. Springer, New York (2011)
436. van den Berg, T.: Calibrating the Ornstein-Uhlenbeck-Vasicek model. Sitmo – Custom Financial Research and Development Services, `www.sitmo.com/article/calibrating-the-ornstein-uhlenbeck-model/` (2011). Retrieved April 20, 2014.
437. van den Bos, A.: Parameter Estimation for Scientists and Engineers. John Wiley & Sons, Hoboken, NJ (2007)
438. van Kampen, N.G.: A power series expansion of the master equation. Can. Chem. Phys. **39**, 551–567 (1961)

439. van Kampen, N.G.: The expansion of the master equation. Adv. Chem. Phys. **34**, 245–309 (1976)
440. van Kampen, N.G.: Remarks on non-markov processes. Brazilian Journal of Physics **28**, 90–96 (1998)
441. van Kampen, N.G.: Stochastic Processes in Physics and Chemistry, third edn. Elsvier, Amsterdam (2007)
442. van Oijen, A.M., Blainey, P.C., Crampton, D.J., Richardson, C.C., Ellenberger, T., Xie, X.S.: Single-moleucles kinetics of λ exconuclease reveal base dependence and dynamic disorder. Science **301**, 1235–1238 (2003)
443. Vasicek, O.: An equlibrium characterization of the term structure. J. Financial Economics **5**, 177–188 (1977)
444. Venn, J.: On the diagrammatic and mechanical representation of propositions and reasonings. The London, Edinburgh, and Diblin Philosophical Magazine and Journal of Science **9**, 1–18 (1880)
445. Venn, J.: Sybolic Logic. MacMillan, London (1881). Second edition, 1984. Reprinted by Lenox Hill Pub. & Dist. Co., 1971
446. Venn, J.: The Logic of Chance. An Essay on the Foundations and Province of the Theory of Probability, with Especial Reference to its Logical Bearings and its Application to Moral and Social Science, and to Statistics, third edn. MacMillan and Co., London (1888)
447. Verhulst, P.: Notice sur la loi que la population pursuit dans son accroisement. Corresp. Math. Phys. **10**, 113–121 (1838)
448. Viswanathan, G.M., Raposo, E.P., da Luz, M.G.E.: Lévy flights and superdiffusion in the context of biological encounters and random searches. Physics of Life Reviews **5**, 133–150 (2008)
449. Vitali, G.: Sul problema della misura dei gruppi di pinti di una retta. Gamberini E. Parmeggiani, Bologna (1905)
450. Vitali, G.: Sui gruppi di punti e sulle funzioni di variabili reali. Atti dell'Accademia delle Science di Torino **43**, 75–92 (1908)
451. Volkenshtein, M.V.: Entropy and Information, *Progress in Mathematical Physics*, vol. 57. Birkhäuser Verlag, Basel, CH (2009). German version: W. Ebeling, Ed. Entropie und Information. Wissenschaftliche Taschenbücher, Band 306, Akademie-Verlag, Berlin 1990. Russian Edition: Nauka Publ., Moscow 1986
452. von Mises, R.: Über Aufteilungs- und Besetzungswahrscheinlichkeiten. Revue de la Faculté des Sciences de l'Université d'Istanbul, N.S. **4**, 145–163 (1938-1939). In German. Reprinted in *Selected Papers of Richard von Mises*, vol.2, American Mathematical Society, 1964, pp. 313-334
453. von Smoluchowski, M.: Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. Annal. Phys. (Leipzig) **21**, 756–780 (1906)
454. Waage, P., Guldberg, C.M.: Studies concerning affinity. J. Chem. Education **63**, 1044–1047 (1986). English translation by Henry I. Abrash
455. Walter, N.G.: Single molecule detection, analysis, and manipulation. In: R.A. Meyers (ed.) Encyclopedia of Analytical Chemistry, pp. 1–10. John Wiley & Sons, Hoboken, NJ (2008)
456. Watson, H.W., Galton, F.: On the probability of the extinction of families. J. Anthropological Institute of Great Britain and Ireland **4**, 138–144 (1875)
457. Weber, N.A.: Dimorphism of the African *oecophylla* worker and an anomaly (*hymenoptera formicidae*). Annals of the Entomological Society of America **39**, 7–10 (1946)
458. Wei, W.W.S.: Time Series Analysis. Univariate and Multivariate Methods. Addison-Wesley Publishing Company, Redwood City, CA (1990)

459. Widengren, J., Mets, Ülo., Rigler, R.: Photodynamic properties of green fluorescent proteins investigated by fluoresecence correlation spectroscopy. Chem. Phys. **250**, 171–186 (1999)
460. Wilheim, T.: The smallest chemical rwaction system with bistability. BMC Systems Biology **3**, e90 (2009)
461. Wilheim, T., Heinrich, R.: Smallest chemical rwaction system with Hopf bifurcation. J.Math. Chem. **17**, 1–14 (1995)
462. Wilkinson, D.J.: Stochastic modeling for quatitative description of heterogeneous biological systems. Nature Rev. Genetics **10**, 122–133 (2009)
463. Wilkinson, D.J.: Stochastic Modelling for Systems Biology, second edn. Chapman & Hall/CRC Press – Taylor and Francis Group, Boca Raton (2012)
464. Williams, D.: Diffusions, Markov Processes and Martingales. Volume 1: Foundations. John Wiley & Sons, Chichester, UK (1979)
465. Winzor, D.J., Jackson, C.M.: Interpretation of the temperature dependence of equilibrium and rate contants. J. Mol. Recognit. **19**, 389–407 (2006)
466. Wolberg, J.: Data Analysis Using the Method of Least Squares. Extracting the Most Information from Experiments. Springer-Verlag, Berlin (2006)
467. Wold, H.: A Study in the Analysis of Time Series, second revised edn. Almqvist and Wiksell Book Co., Uppsala, SE (1954). With an appendix on *Recent Developments in Time Series Analysis* by Peter Whittle.
468. Wright, S.: Evolution in Mendelian populations. Genetics **16**, 97–159 (1931)
469. Wright, S.: The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: D.F. Jones (ed.) Int. Proceedings of the Sixth International Congress on Genetics, vol. 1, pp. 356–366. Brooklyn Botanic Garden, Ithaca, NY (1932)
470. Yang, Y., Rathinam, M.: Tau leaping of stiff stochastical chemical systems via local central limit approximation. J. Comp. Phys. **242**, 581–606 (2013)
471. Yashonath, S.: Relaxation time of chemical reactions from network thermodynamics. J. Phys. Chem. **85**, 1808–1810 (1981)
472. Zhabotinsky, A.M.: A history of chemical oscillations and waves. Chaos **1**, 379–386 (1991)
473. Zhang, W.K., Zhang, X.: Single molecule mechanochemistry of macromolecules. Prog. Polym. Sci. **28**, 1271–1295 (2003)
474. Zwillinger, D.: Handbook of Differential Equations, third edn. Academic Press, San Diego, CA (1998)

Index

Author Index

Notation

Mathematical symbols:

symbol	usage	interpretation
$\{\ \}$	$\{A, B, \ldots\}$	a set consisting of elements $A, B, \ldots$
$\emptyset$		empty set
Ω		entire sample space, universe
$\mid$	$\{A \mid C(A)\}$	elements of A, which fulfil condition $C(A)$
:	$\{A : C(A)\}$	elements of A, which fulfil condition $C(A)$
$\circ$	$T_2 \circ T_1(\cdot)$	composition, sequential operation on $(\cdot)$
$*$	$f(t) * g(t)$	convolution, $\big(f * g\big)(t) = \int_{-\infty}^{\infty} f(\tau)\, g(t-\tau)\mathrm{d}\tau$
$\doteq$	$a \doteq b$	definition
$\xrightarrow{d}$	$\lim\limits_{n\to\infty} \langle f(\mathcal{X}_n)\rangle \xrightarrow{d} \langle f(\mathcal{X})\rangle$	convergence in distribution
$\otimes$	$\mathbf{e}_1 \otimes \mathbf{e}_2 \otimes \mathbf{e}_3$	Cartesian product used for 3D-space
$\bigotimes_{k=1}^{n}$	$\mathbf{e}_1 \otimes \ldots \otimes \mathbf{e}_n$	Cartesian product used for nD-space
$\rightarrow$		
$\Rightarrow$		
log		logarithm in general and logarithm to base 10
ln		natural logarithm or logarithm to base e
ld		logarithm to base 2

Selected mathematical functions:

symbol	usage	interpretation
π	$\pi(\alpha)$	Poisson distribution
$\mathcal{N}$	$\mathcal{N}(\mu, \sigma^2)$	normal distribution
$\mathcal{U}$	$\mathcal{U}(\,\Omega)$	uniform distribution over sample space Ω

Vectors and matrices:

$\mathbf{e}_k$	$(0,\ldots,1,\ldots,0)$	unit vector in the direction of the k-th coordinate

Number systems:

natural numbers	$\mathbb{N}$	$\{0,1,2,3,\cdots\}$
	$\mathbb{N}_{>0}$	$\{1,2,3,\cdots\}$
integers	$\mathbb{Z}$	$\{\cdots,-3,-2,-1,0,1,2,3,\cdots\}$
	$\mathbb{Z}_{>0}=\mathbb{N}_{>0}$	$\{1,2,3,\cdots\}$
	$\mathbb{Z}_{\geq 0}=\mathbb{N}$	$\{0,1,2,3,\cdots\}$
	$\mathbb{Z}_{<0}$	$\{-1,-2,-3,\cdots\}$
	$\mathbb{Z}_{\leq 0}$	$\{0,-1,-2,-3,\cdots\}$
rational numbers	$\mathbb{Q}$	$\{\frac{m}{n}\,\vert\,(m,n)\in\mathbb{Z}\wedge n\neq 0\}$
real numbers	$\mathbb{R}$	$\{x\,\vert\,x$ is rational or irrational$\}$
complex numbers	$\mathbb{C}$	$\{z=a+b\,\mathrm{i}\,\vert\,(a,b)\in\mathbb{R}\,,\,\mathrm{i}=\sqrt{-1}\}$

Variables:

	symbols	functions
discrete variables	$i,j,k,(l),n,m,\ldots,T$	$f_k\,,\,P_k(t)\,,\,\ldots$
continuous variables	$x,y,z,\ldots,r,s,t$	$f(x)\,,\,P(x,t)\,,\,\ldots$
random variables	$\mathcal{A},\mathcal{B},\ldots,\mathcal{X},\mathcal{Y},\mathcal{Z}$	
discrete	$(\mathcal{K},\mathcal{M},\mathcal{N},\ldots)\in\mathbb{N}$	$f(\mathcal{N})=f_n\,,\ldots$
continuous	$(\mathcal{X},\mathcal{Y},\mathcal{Z},\ldots)\in\mathbb{R}$	$f(\mathcal{X})=f(x)\,,\ldots$

Quantities:

numbers of particles of $\mathsf{A},\mathsf{B},\ldots,$	$N_{\mathsf{A}},N_{\mathsf{B}},\ldots,$
Avogadro constant	$N_{\mathrm{L}}=6.02214179\times 10^{23}\,\mathrm{mol}^{-1}$

Chemical and biological species:	variables
specific entities: $\mathsf{A}, \mathsf{B}, \ldots,$	$a = [\mathsf{A}]$, $b = [\mathsf{B}]$, $\ldots,$
autocatalytic entities: $\mathsf{X}, \mathsf{Y}, \ldots,$	$x = [\mathsf{X}]$, $y = [\mathsf{Y}]$, $\ldots,$
unspecific entities: $\mathsf{X}_1, \mathsf{X}_2, \ldots,$	$x_1 = [\mathsf{X}_1]$, $x_2 = [\mathsf{X}_2]$, $\ldots,$

Thermodynamics:	symbol	reference state
energy:		
entropy:		
enthalpy:		
Gibbs free energy:	ΔG_0	
chemical potential:		

Kinetics:	symbol	usage
rate function:	$v_{\boldsymbol{\mu}}(\mathbf{n}) = \gamma_{\boldsymbol{\mu}} h_{\boldsymbol{\mu}}(\mathbf{n})$	reaction $\mathbf{R}_{\boldsymbol{\mu}}$
rate parameter:	$\gamma_{\boldsymbol{\mu}}$	reaction $\mathbf{R}_{\boldsymbol{\mu}}$
	$k_{\boldsymbol{\mu}}, l_{\boldsymbol{\mu}}$	deterministic kinetics
	$\kappa_{\boldsymbol{\mu}}, \lambda_{\boldsymbol{\mu}}$	stochastic kinetics
stoichiometric factor:	$h_{\boldsymbol{\mu}}(\mathbf{n}) = \prod_{j=1}^{M} n_j^{\nu_{j\boldsymbol{\mu}}}$	reaction $\mathbf{R}_{\boldsymbol{\mu}}$

usage of $\cdot$ and $\times$
numbers and spaces: $\mathbb{N}$, $\mathbb{N}_{>0}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{R}^n$
logical operators: $\forall$, $\rightarrow$, $\Longrightarrow$
scaling parameter: σ
support: 'supp', for example in $\bigcup_{\text{supp}}$
definition: ':='
proportional to: '$\propto$'
vectors and matrices: transposition $^{\mathrm{t}}$'
unit matrix: $\mathbb{I}$
linear span: 'span': $\operatorname{span}(S) = \left\{\sum_{i=1}^{k} \lambda_i u_i | k \in \mathbb{N}, u_i \in S, \lambda_i \in \mathbf{K}\right\}$, S is a finite subset of a vector space $\mathcal{U}$ over a field $\mathbf{K}$.
concentration vectors: $\mathbf{x} = (a, b, \ldots) = ([\mathbf{A}], [\mathbf{B}], \ldots)$
random vector: $\vec{\mathcal{X}} = (\mathcal{X}_1, \ldots, \mathcal{X}_M)$,
discrete realization: $\mathbf{n} = (\mathrm{n}_1, \ldots, \mathrm{n}_M)$,
continuous realization: $\mathbf{x} = (x_1, \ldots, x_M)$
reaction rate: $v\big(\mathbf{x}(t)\big)$, $\mathbf{v}\big(\mathbf{x}(t)\big)$

Units and conversion factors
volume: liter $1\,\mathrm{l} = 1\,\mathrm{dm}^3 = 0.001\,\mathrm{m}^3$
particle numbers: N
Arbitrary units: $[l.u.]$ ... length unit

Frequently used transforms:

Fourier transform: $\mathcal{F}\big(f(x)\big)(k) = \tilde{f}(k) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} f(x)\, e^{\mathrm{i}\, k\, x}\, \mathrm{d}x$

Laplace transform: $\mathcal{L}\big(f(t)\big)(s) = \hat{f}(s) = \int\limits_{0}^{\infty} e^{-st}\, f(t)\ \mathrm{dt}$

Fourier-Laplace transform:
$\mathcal{L}\mathcal{F}\big(f(x,t)\big)(k,s) = \hat{\tilde{f}}(k,s) = \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{\infty} \int\limits_{-\infty}^{\infty} e^{-st}\, e^{\mathrm{i}\, k\, x}\, f(x,t)\, \mathrm{d}x\ \mathrm{dt}$