Evolutionary Dynamics

A Special Course on Dynamical Systems and Evolutionary Processes for Physicists, Chemists, and Biologists

Summer Term 2014

Version of February 24, 2014

Peter Schuster

Institut für Theoretische Chemie der Universität Wien

Währingerstraße 17, A-1090 Wien, Austria

Phone: +43 1 4277 52736 , eFax: +43 1 4277 852736

E-Mail: pks@tbi.univie.ac.at

Internet: http://www.tbi.univie.ac.at/~pks

Preface

The current text contains notes that were prepared first for a course on 'Evolutionary Dynamics' held at Vienna University in the summer term 2014. No claim is made that the text is free of errors. The course is addressed to students of physics, chemistry, biology, biochemistry, molecular biology, mathematical and theoretical biology, bioinformatics, and systems biology with particular interests in evolutionary phenomena. Evolution although in the heart of biology as expressed in Theodosius Dobzhansky famous quote,

"nothing in biology makes sense except in the light of evolution",

is a truly interdisciplinary subject and hence the course will contain elements from various disciplines, mainly from mathematics, in particular dynamical systems theory and stochastic processes, computer science, chemical kinetics, molecular biology, and evolutionary biology. Considerable usage of mathematical language and analytical tools is indispensable, but we have consciously avoided to dwell upon deeper and more formal mathematical topics.

Evolution has been shifted into the center of biological thinking through Charles Darwin's centennial book 'On the Origin of Species' [45]. Gregor Mendel's discovery of genetics [206] was the second milestone of evolutionary biology but it remained largely ignored for almost forty years before it became first an alternative concept to selection. Biologists were split into two camps, the *selectionists* believing in continuity in evolution and the *geneticists*, who insisted in the discreteness of change in the form of mutation (An account of the historic development of mutation as an ides is found in the recent publication [34]. The unification of two concepts was first achieved on the level of a mathematical theory through population genetics [92, 318] developed by the three great scholars Ronald Fisher, J.B.S. Haldane, and Sewall Wright. Still it took twenty more years before the *synthetic theory* of evolution had been completed [202]. Almost all attempts of biologists to understand evolution were and most of them still are completely free of quantitative or mathematical thinking. The two famous exceptions are Mendelian genetics and populations genetics. It is impossible, however, to model or understand dynamics without quantitative description. Only recently and mainly because of the true flood of hitherto unaccessible data the desire for a new and quantitative theoretical biology has been articulated [26, 27]. We shall focus in this course on dynamical models of evolutionary

processes, which are rooted in physics, chemistry, and molecular biology. On the other hand, any useful theory in biology has to be grounded on a solid experimental basis. Most experimental data on evolution at the molecular level are currently focussing on genomes and accordingly, sequence comparisons and reconstruction of phylogenetic trees are a topic of primary interest [232]. The fast, almost explosive development of molecular life sciences has reshaped the theory of evolution [276]. RNA has been considered as a rather uninteresting molecule until the discovery of RNA catalysis by Thomas Cech and Sidney Altman in the nineteen eighties, nowadays RNA is understood as an important regulator of gene activity [9], and we have definitely not come near to the end of the exciting RNA story.

This series of lectures will concentrate on principles rather than technical details. At the same time it will be necessary to elaborate tools that allow to treat real problems. The tools required for the analysis of dynamical systems are described, for example, in the two monographs [143, 144]. For stochastic processes we shall follow the approach taken in the book [107] and presented in the course of the Summer term 2011 [36, 251]. Some of the stochastic models in evolution presented here are described in the excellent review [22]. Analytical results on evolutionary processes are rare and thus it will be unavoidable to deal also with approximation methods and numerical techniques that are able to produce results through computer calculations (see, for example, the article [112, 113, 115, 117]). The applicability of simulations to real problems depends critically on population sizes that can by handled. Present day computers can readily deal with 10^6 to 10^7 particles, which is commonly not enough for chemical reactions but sufficient for most biological problems and accordingly the sections dealing with practical examples will contain more biological than chemical problems. A number of text books have been used in the preparation of this text in addition to the web encyclopedia *Wikipedia*. In molecular biology, molecular genetics, and population genetics these texts were [5, 125, 130, 136]

The major goal of this text is to avoid distraction of the audience by taking notes and to facilitate understanding of subjects that are quite sophisticated at least in parts. At the same time the text allows for a repetition of the major issues of the course. Accordingly, an attempt was made in preparing a useful and comprehensive list of references. To study the literature in detail is recommended to every serious scholar who wants to progress towards a deeper understanding of this rather demanding discipline.

Peter Schuster Wien, February 2014.

1. Darwin's principle in mathematical language

Charles Darwin's principle of natural selection is a powerful abstraction from observations, which provides insight into the basic mechanism giving rise to changing species. Species or populations don't multiply but individuals do, either directly in asexual species, like viruses, bacteria or protists, or in sexual species through pairings of individuals with opposite sex. Variability of individuals in populations is an empirical fact that can be seen easily in everyday life. Within populations the variants are subjected to natural selection and those having more progeny prevail in future generations. The power of Darwin's abstraction lies in the fact that neither the shape and the structure of individuals nor the mechanism of inheritance are relevant for selection unless they have an impact on the number of offspring. Otherwise Darwin's approach had been doomed to fail since his imagination of inheritance was incorrect. Indeed Darwin's principle holds simultaneously for highly developed organisms, for primitive unicellular species like bacteria, for viruses and even for reproducing molecules in cell-free assays.

Molecular biology provided a powerful possibility to study evolution in its simplest form outside biology: Replicating ribonucleic acid molecules (RNA) in cell-free assays [268] play natural selection in its purest form: In the test tube, evolution, selection, and optimization are liberated from all unnecessary complex features, from obscuring details, and from unimportant accessories. Hence, *in vitro* evolution can be studied by the methods of chemical kinetics. The parameters determining the "fitness of molecules" are replication rate parameters, binding constants, and other measurable quantities, which can be determined independently of *in vitro* evolution experiments, and constitute an alternative access to the determination of the outcome of selection. Thereby "survival of the fittest" is unambiguously freed from the reproach of being the mere tautology of "survival of the survivor". In addition, *in vitro* selection turned out to be extremely useful for the synthesis

of molecules that are tailored for predefined purposes. A new area of applications called evolutionary biotechnology branched off evolution in the test tube. Examples for evolutionary design of molecules are [166, 176] for nucleic acids, [25, 161] for proteins, and [316] for small organic molecules.

The chapter starts by mentioning a few examples of biological applications of mathematics before Darwin (section 1.1), then we derive and analyze an ODE describing simple selection with asexual species (section 1.2), and consider the effects of variable population size (section 1.3). The next subsection 1.4 analyzes optimization in the Darwinian sense, and eventually we consider generic properties of typical growth functions (section 1.5).

1.1 Counting and modeling before Darwin

The first mathematical model that seems to be relevant for evolution was conceived by the medieval mathematician Leonardo Pisano also known as Fibonacci. His famous book *Liber abaci* has been finished and published in the year 1202 and was translated into modern English eight years ago [264]. Among several other important contributions to mathematics in Europe Fibonacci discusses a model of rabbit multiplication in *Liber abaci*. Couples of rabbits reproduce and produce young couples of rabbits according to the following rules:

(i) Every adult couple has a progeny of one young couple per month,

(ii) a young couple grows to adulthood within the first month and accordingly begins producing offspring in the second months,

(iii) rabbits live forever, and

(iv) the number of rabbit couples is updated every month.

The model starts with one young couple (**1**), nothing happens during maturation of couple **1** in the first month and we have still one couple in the second month. In the third month, eventually, a young couple (**2**) is born and the number of couples increases to two. In the fourth month couple **1** produces a new couple (**3**) whereas couple **2** is growing to adulthood, and

we have three couples now. Further rabbit counting yields the Fibonacci sequence:[1]

month	0	1	2	3	4	5	6	7	8	9	...
# couples	0	1	1	2	3	5	8	13	21	34	...

It is straightforward to derive a recursion for the rabbit count. The number of couples in month $(n+1)$, f_{n+1}, is the sum of two terms: The number of couples in month n, because rabbits don't die, plus the number of young couples that is identical to the number of couples in month $(n-1)$:

$$f_{n+1} = f_{n-1} + f_n \quad \text{with} \quad f_0 = 0 \;\; \text{and} \;\; f_1 = 1\,. \tag{1.1}$$

With increasing n the ratio of two subsequent Fibonacci numbers converges to the golden ratio, $f_{k+1}/f_k = (1+\sqrt{5})/2$ (For a comprehensive discussion of the Fibonacci sequence and its properties see [124, pp.290-301] or, e.g., [61]).

In order to proof this convergence we make use of a matrix representation of the Fibonacci model:

$$\mathrm{F}^n \begin{pmatrix} f_0 \\ f_1 \end{pmatrix} = \begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} \;\; \text{with} \;\; \mathrm{F} = \begin{pmatrix} f_0 & f_1 \\ f_1 & f_2 \end{pmatrix} \;\; \text{and} \;\; \mathrm{F}^n = \begin{pmatrix} f_{n-1} & f_n \\ f_n & f_{n+1} \end{pmatrix}.$$

The matrix representation transforms the recursion into an expression that allows for direct computation of the elements of the Fibonacci sequence.

$$f_n = \begin{pmatrix} 1 & 0 \end{pmatrix} \mathrm{F}^n \begin{pmatrix} f_0 \\ f_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} f_{n-1} & f_n \\ f_n & f_{n+1} \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \end{pmatrix}. \tag{1.2}$$

Theorem 1.1 (Fibonacci convergence)**.** *With increasing n the Fibonacci sequence converges to a geometric progression with the golden ratio as factor, $q = (1+\sqrt{5})/2$.*

Proof. The matrix F is diagonalized by the transformation $T^{-1} \cdot \mathrm{F} \cdot \mathrm{T} = \mathrm{D}$ with $\mathrm{D} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$. The two eigenvalues of F are: $\lambda_{1,2} = (1 \pm \sqrt{5})/2$. Since

[1] According to Parmanand Singh [265] the Fibonacci numbers were invented earlier in India and used for the solution of various problems (See also Donald Knuth [178]).

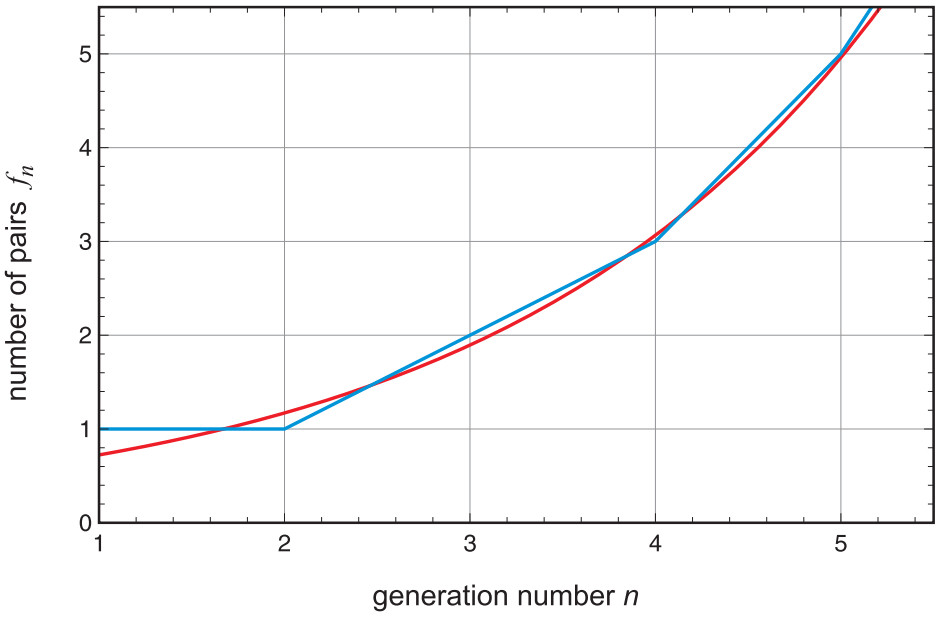


Figure 1.1: **Fibonacci series and geometric progression**. The Fibonacci series (1.1) (blue) is compared with the geometric progression $g_n = q^n/\sqrt{5}$ with $q = (1+\sqrt{5})/2$ (red). The Fibonacci series oscillates around the geometric progression with decreasing amplitude and converges asymptotically to it.

F is a symmetric matrix the $\mathbb{L}^2$-normalized eigenvectors of F, $(\boldsymbol{e}_1, \boldsymbol{e}_2) = \mathrm{T}$, form an orthonormal set,

$$T = \begin{pmatrix} \frac{1}{\sqrt{1+\lambda_1^2}} & \frac{1}{\sqrt{1+\lambda_2^2}} \\ \frac{\lambda_1}{\sqrt{1+\lambda_1^2}} & \frac{\lambda_2}{\sqrt{1+\lambda_2^2}} \end{pmatrix} \quad \text{and} \quad T \cdot T' = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

with T' being the transposed matrix, and $\mathrm{T}^{-1} = \mathrm{T}'$. Computation of the n-th power of matrix F yields

$$\mathrm{F}^n = T \cdot \mathrm{D}^n \cdot \mathrm{T}' = \mathrm{T} \cdot \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} \cdot \mathrm{T}' = \frac{1}{\sqrt{5}} \begin{pmatrix} \lambda_1^{n-1} - \lambda_2^{n-1} & \lambda_1^n - \lambda_2^n \\ \lambda_1^n - \lambda_2^n & \lambda_1^{n+1} - \lambda_2^{n+1} \end{pmatrix} ,$$

from which the expression for f_n is obtained by comparison with (1.2)

$$f_n = \frac{1}{\sqrt{5}}(\lambda_1^n - \lambda_2^n) \; . \tag{1.3}$$

Because $\lambda_1 > \lambda_2$ the ratio converges to zero: $\lim_{n\to\infty} \lambda_2^n/\lambda_1^n = 0$, and the Fibonacci sequence is approximated well by a geometric progression $f_n \approx g_n = \frac{1}{\sqrt{5}}\, q^n$ with $q = (1+\sqrt{5})/2$. □

Since λ_2 is negative the Fibonacci sequence alternates around the geometric progression. Expression (1.3) is commonly attributed to the French mathematician Jacques Binet [21] and named after him. As outlined in ref. [124, p.299] the formula has been derived already hundred years before by the great Swiss mathematician Leonhard Euler [80] but was forgotten and rediscovered.

Thomas Robert Malthus was the first who articulated the ecological and economic problem of population growth following a geometric progression [193]: Animal or human populations like every system capable of reproduction grow like a geometric progression provided unlimited resources are available. The resources, however, are either constant or grow – as Malthus assumes – according to an arithmetic progression if human endeavor is involved. The production of nutrition, says Malthus, is proportional to the land that is exploitable for agriculture and the gain in the area of fields will be a constant in time – the increase will be the same every year. An inevitable result of Malthus' vision of the world is the pessimistic view that populations will grow until the majority of individuals will die premature of malnutrition and hunger. Malthus could not foresee the *green revolutions* but he was also unaware that population growth can be faster than exponential – sufficient nutrition for the entire human population is still a problem. Charles Darwin and his younger contemporary Alfred Russel Wallace were strongly influenced by Robert Malthus and took form population theory that in the wild, where birth control does not exist and individuals fight for food, the major fraction of of progeny will die before they reach the age of reproduction and only the strongest will have a chance to multiply.

Leonhard Euler introduced the notions of the exponential function in the middle of the eighteenth century [81] and set the stage for modeling growing populations by means of ordinary differential equations (ODEs). The growth

rate is proportional to the number of individuals or the population size N

$$\frac{\mathrm{d}N}{\mathrm{dt}} \;=\; r\,N \;, \tag{1.4}$$

where the parameter r is commonly called Malthus or growth parameter. Straightforward integration yields:

$$\int_{N(0)}^{N(t)} \frac{\mathrm{d}N}{N} \;=\; \int_0^t \mathrm{dt} \quad \text{and} \quad N(t) = N_0 \exp(rt) \quad \text{with} \quad N_0 = N(0) \;. \tag{1.5}$$

Simple reproduction results in exponential growth of a population with $N(t)$ individuals.

Presumably not known to Darwin, the mathematician Jean François Verhulst complemented the concept of exponential growth by the introduction of finite resources [292–294]. The Verhulst equation is of the form[2]

$$\frac{\mathrm{d}N}{\mathrm{dt}} = rN\left(1-\frac{N}{K}\right) \;, \tag{1.6}$$

where $N(t)$ again denotes the number of individuals of a species **X**, and K is the *carrying capacity* of the ecological niche or the ecosystem. Equ. (1.6) can be integrated by means of partial fractions ($\gamma = 1/K$):

$$\int_{N_0}^{N(t)} \frac{\mathrm{d}N}{N(1-\gamma N)} \;=\; \int_{N_0}^{N(t)} \frac{\mathrm{d}N}{N} + \int_{N_0}^{N(t)} \frac{\gamma\,\mathrm{d}N}{1-\gamma N} \;,$$

and the following solution is obtained

$$N(t) = N_0\,\frac{K}{N_0 + \bigl(K-N_0\bigr)\exp(-rt)} \;. \tag{1.7}$$

Apart from the initial condition N_0, the number of individuals **X** at time $t = 0$, the logistic equation has two parameters: (i) the Malthusian parameter or the growth rate r and (ii) the carrying capacity K of the ecological niche or the ecosystem. A population of size N_0 grows exponentially at short times: $N(t) \approx N_0 \exp(rt)$ for $K \gg N_0$ and t sufficiently small. For long

[2]The Verhulst equation is also called logistic equation and its discrete analogue is the logistic map, a standard model to demonstrate the occurrence of deterministic chaos in a simple system. The name *logistic equation* was coined by Verhulst himself in 1845.

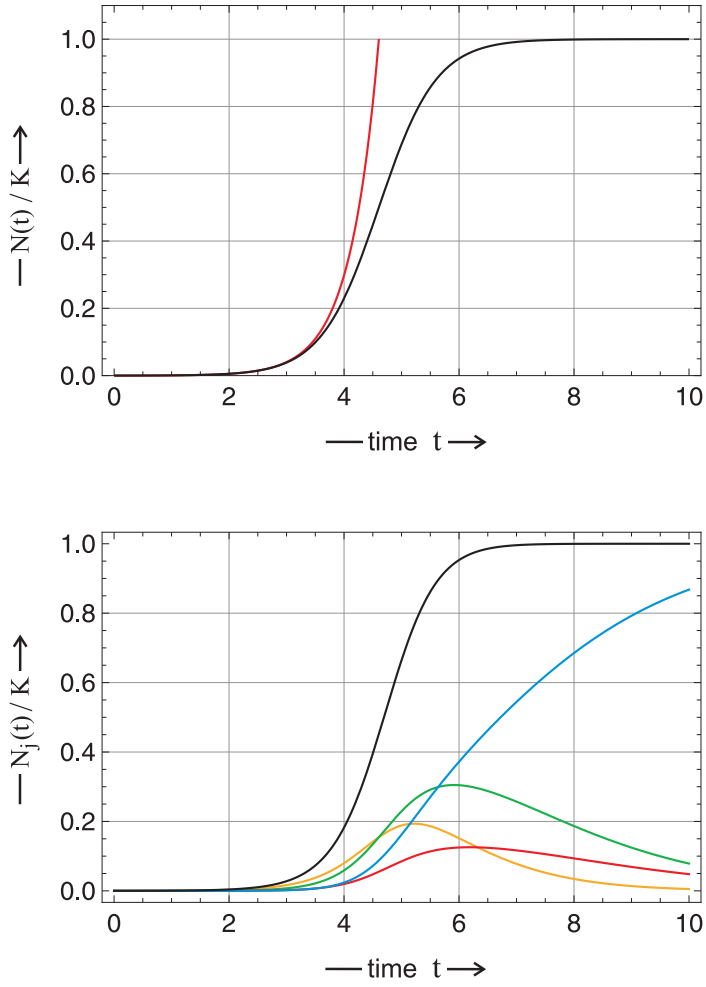


Figure 1.2: **Solution curves of the logistic equations** (1.7,1.13). Upper plot: The black curve illustrates growth in population size from a single individual to a population at the carrying capacity of the ecosystem. The red curve represents the results for unlimited exponential growth, $N(t) = N(0)\exp(rt)$. Parameters: $r = 2$, $N(0) = 1$, and $K = 10\,000$. Lower plot: Growth and internal selection is illustrated in a population with four variants. Color code: C black, N_1 yellow, N_2 green, N_3 red, N_4 blue. Parameters: fitness values $f_j = (1.75, 2.25, 2.35, 2.80)$, $N_j(0) = (0.8888, 0.0888,\ 0.0020, 0.0004)$, $K = 10\,000$. The parameters were adjusted such that the curves for the total populations size $N(t)$ coincide (almost) in both plots.

times the population size approaches the carrying capacity asymptotically: $\lim_{t\to\infty} N(t) = K$.

The two parameters r and K are taken as criteria to distinguish different evolutionary strategies: Species that are r-selected exploit ecological niches with low density, produce a large number of offspring each of which has a low probability to survive, whereas K-selected species are strongly competing in crowded niches and invest heavily in few offspring that have a high probability of survival to adulthood. The two cases, r- and K-selection, are the extreme situations of a continuum of mixed selection strategies. In the real world the r-selection strategy is an appropriate adaptation to fast changing environments, whereas K-selection pays in slowly varying or constant environments.

1.2 The selection equation

The logistic equation can be interpreted differently and this is useful is the forthcoming analysis: In the second term — $-(N/K)\,rN$ — the expression rN/K is identified with a constraint for limiting growth: $rN/K \equiv \phi(t)$,

$$\frac{\mathrm{d}N}{\mathrm{dt}} = N\Big(r - \phi(t)\Big)\ , \tag{1.6'}$$

The introduction of $\phi(t)$ gives room for other interpretations of constraints than carrying capacities of ecosystems. For example, $\phi(t)$ may be a dilution flux in laboratory experiments on evolution in flow reactors [234, pp.21-27]. Equ. (1.6') falls into the class of *replicator equations*, $dx/dt = x\,F(x)$ [253], which describe the time development of the concentrations of *replicators* $\mathbf{X}$.

Equ. (1.6') can be used now for the derivation of a selection equation in the spirit of Darwin's theory. The single species $\mathbf{X}$ is replaced by several variants forming a population, $\Upsilon = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$; in the language of chemical kinetics competition and selection are readily cast into a reaction mechanism consisting of n independent, simple replication reactions:

$$(\mathbf{A}) + \mathbf{X}_j \xrightarrow{f_j} 2\,\mathbf{X}_j\ , \quad j = 1, 2, \ldots, n\ . \tag{1.8}$$

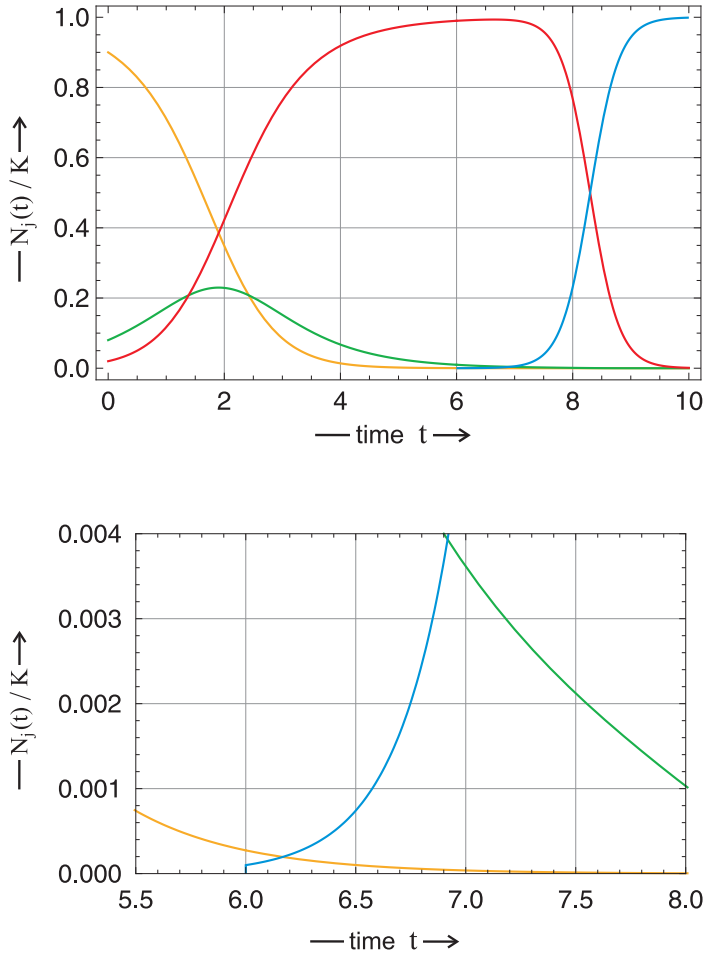


Figure 1.3: **Solution curve of the selection equation** (1.13). The system is studied at constant maximal population size, $N = K$, and relative concentrations are applied: $x_j = N_j/K$. The plots represent calculated changes of the variant distributions with time. The upper plot shows selection among three species $\mathbf{X}_1$ (yellow), $\mathbf{X}_2$ (green), and $\mathbf{X}_3$ (red), and then the appearance of a fourth, fitter variant $\mathbf{X}_4$ (blue) at time $t = 6$, which takes over and becomes selected thereafter. The lower plot presents an enlargement of the upper plot around the point of spontaneous creation of the fourth species ($\mathbf{X}_4$). Parameters: fitness values $f_j = (1, 2, 3, 7)$; $x_j(0) = (0.9, 0.08, 0.02, 0)$ and $x_4(6) = 0.0001$.

The symbol $\mathbf{A}$ denotes the material from which $\mathbf{X}_j$ is synthesized (It is put in parentheses, because we assume that it is present in access and its concentration is constant therefore). The numbers of individuals of the variants are denoted by $N_j(t)$, or in vector notation $\boldsymbol{N}(t) = \big(N_1(t), N_2(t), \ldots, N_n(t)\big)$ with $\sum_{i=1}^{n} N_i(t) = C(t)$. A common carrying capacity is defined for all n variants:

$$\lim_{t\to\infty} \sum_{i=1}^{n} N_i(t) = \lim_{t\to\infty} C(t) = K \ .$$

The Malthus parameters are given here by the fitness values $f_1, f_2, \ldots, f_n$, respectively. For individual species the differential equations take on the form

$$\begin{aligned} \frac{\mathrm{d}N_j}{\mathrm{dt}} &= N_j \left(f_j - \frac{C}{K}\phi(t) \right) ; \ j = 1, 2, \ldots, n \quad \text{with} \\ \phi(t) &= \frac{1}{C} \sum_{i=1}^{n} f_i N_i(t) \end{aligned} \tag{1.9}$$

being the mean fitness of the population. Summation over all species yields a differential equation for the total population size

$$\frac{\mathrm{d}C}{\mathrm{dt}} = C \left(1 - \frac{C}{K} \right) \phi(t) \ . \tag{1.10}$$

Stability analysis is straightforward: From $\mathrm{d}C/\,\mathrm{dt} = 0$ follow two stationary states of equ. (1.10): (i) $\bar{C} = 0$ and (ii) $\bar{C} = K$.[3] For conventional stability analysis we calculate the (1×1) Jacobian and obtain for the eigenvalue

$$\lambda = \frac{\partial\big(\mathrm{d}C/\,\mathrm{dt}\big)}{\partial C} = \phi(t) - \frac{C}{K} \left(2\phi(t) - K\frac{\partial\phi}{\partial C} \right) - \frac{C^2}{K}\frac{\partial\phi}{\partial C} \ .$$

Insertion of the stationary values yields $\lambda_{(i)} = \phi > 0$ and $\lambda_{(ii)} = -\phi < 0$, state (i) is unstable and state (ii) is asymptotically stable. The total population size converges to the value of the carrying capacity, $\lim_{t\to\infty} C(t) = \bar{C} = K$.

[3] There is also a third stationary state defined by $\phi = 0$. For strictly positive fitness values, $f_i > 0 \,\forall\, i = 1, 2, \ldots, n$, this condition can only be fulfilled by $N_i = 0 \,\forall\, i = 1, 2, \ldots, n$, which is identical to state (i). If some f_i values are zero – corresponding to lethal variants – the respective variables vanish in the infinite time limit because of $\mathrm{d}N_i/\,\mathrm{dt} = -\phi(t)\, N_i$ with $\phi(t) > 0$.

Equ. (1.10) can be solved exactly yielding thereby an expression that contains integration of the constraint $\phi(t)$:

$$C(t) = C(0)\frac{K}{C(0) + \big(K - C(0)\big)\exp(-\Phi)} \quad \text{with} \quad \Phi = \int_0^t \phi(\tau)\mathrm{d}\tau \ ,$$

where $C(0)$ is the population size at time $t = 0$. The function $\Phi(t)$ depends on the distribution of fitness values within the population and its time course. For $f_1 = f_2 = \ldots = f_n = r$ the integral yields $\Phi = rt$ and we retain equ. (1.7). In the long time limit Φ grows to infinity and $C(t)$ converges to the carrying capacity K.

At constant population size $C = \bar{C} = K$ equ. (1.9) becomes simpler

$$\frac{\mathrm{d}N_j}{\mathrm{dt}} = N_j\Big(f_j - \phi(t)\Big)\,; \quad j = 1, 2, \ldots, n\,. \tag{1.9'}$$

and can be solved exactly by means of the integrating factor transformation [329, p. 322ff.]:

$$Z_j(t) = N_j(t)\exp\left(\int_0^t \phi(\tau)\,\mathrm{d}\tau\right) \ . \tag{1.11}$$

Insertion into equ. (1.9') yields

$$\begin{aligned}
\frac{\mathrm{d}N_j}{\mathrm{dt}} &= \frac{\mathrm{d}Z_j}{\mathrm{dt}}\exp\left(\int_0^t -\phi(\tau)\,\mathrm{d}\tau\right) - Z_j\exp\left(\int_0^t -\phi(\tau)\,\mathrm{d}\tau\right)\phi(t) = \\
&= Z_j\exp\left(\int_0^t -\phi(\tau), \mathrm{d}\tau\right)\Big(f_j - \phi(t)\Big)\ , \\
\frac{\mathrm{d}Z_j}{\mathrm{dt}} &= Z_j\ ,\ j = 1, 2, \ldots, n \quad \text{or} \quad \mathrm{d}\boldsymbol{Z}/\,\mathrm{dt} = \mathrm{F}\cdot\boldsymbol{Z}\ ,
\end{aligned} \tag{1.12}$$

where F is a diagonal matrix containing the fitness values f_j $(j = 1, 2, \ldots, n)$ as elements. Using the trivial equality $Z_j(0) = N_j(0)$ we obtain for the individual genotypes:

$$N_j(t) = N_j(0)\exp(f_j t)\,\frac{C}{\sum_{i=1}^{n} N_i(0)\exp(f_i t)}; \ j = 1, 2, \ldots, n\ . \tag{1.13}$$

Equ. (1.13) encapsulates Darwinian selection and optimization of fitness in populations that will be discussed in detail in section 1.4.

The use of normalized or internal variables $x_j = N_j/C$ provides certain advantages and we shall use them whenever we are dealing with constant population size. The ODE is of the form

$$\begin{aligned} \frac{\mathrm{d}x_j}{\mathrm{dt}} &= f_j x_j - x_j\phi(t) = x_j\Big(f_j - \phi(t)\Big);\ j = 1, 2, \ldots, n \quad \text{with} \\ \phi(t) &= \sum_{i=1}^{n} f_i x_i \ , \end{aligned} \tag{1.14}$$

the solution is trivially the same as in case of equ.(1.9):

$$x_j(t) = \frac{x_j(0)\exp(f_j t)}{\sum_{i=1}^{n} x_i(0)\exp(f_i t)};\ j = 1, 2, \ldots, n \ . \tag{1.15}$$

The use of normalized variables, $\sum_{i=1}^{n} x_i = 1$, defines the unit simplex, $\mathbb{S}_n^{(1)} = \{0 \leq x_i \leq 1 \forall\, i = 1, \ldots, n \wedge \sum_{i=1}^{n} x_i = 1\}$, as the physically accessible domain that fulfils the conservation relation. All boundaries of the simplex — corners, edges, faces, etc. — are invariant sets, since $x_j = 0 \Rightarrow \mathrm{d}x_j/\,\mathrm{dt} = 0$ by equ. (1.14).

Asymptotic stability of the simplex follows from the stability analysis of equ. (1.10) and implies that all solution curves converge to the unit simplex from every initial condition, $\lim_{t\to\infty}\Big(\sum_{i=1}^{n} x_i(t)\Big) = 1$. In other words, starting with any initial value $C(0) \neq 1$ the population approaches the unit simplex. When it starts on $\mathbb{S}_n$ it stays there and in presence of fluctuations it will return to the invariant manifold. As long as the population is finite, $0 < C < +\infty$, and since $N_j(t) = x_j(t) \cdot C(t)$, we can restrict population dynamics to the unit simplex without loosing generality and characterize the state of a population at time t by the vector $\boldsymbol{x}(t)$ which fulfils the $\mathbb{L}^{(1)}$ norm $\sum_{i=1}^{n} x_i(t) = 1$ (as an example see fig. 1.4). In the next section 1.3 we shall consider variable $C(t)$ explicitly.

1.3 Variable population size

Now we shall show that the solution of equ. (1.9) describes the internal equilibration for constant and variable population sizes as long as the population does neither explode nor die out, $0 < C(t) < +\infty$ [71]. The validity of theorem 1.2 as will be shown below is not restricted to constant

fitness values f_j and hence we can replace them by general growth functions $G_j(N_1, \ldots, N_n) = G_j(\boldsymbol{N})$ or fitness functions $F_j(\boldsymbol{N})$ with $G_j(\boldsymbol{N}) = F_j(\boldsymbol{N})N_j$ in the special case of replicator equations [253]: $\mathrm{d}N_j/\,\mathrm{dt} = N_j\big(F_j(\boldsymbol{N}) - \Psi(t)\big)$ where $\Psi(t)$ comprises both, variable total concentration and constraint.

Time dependence of the conditions in the ecosystem can be introduced in two ways: (i) variable carrying capacity, $K(t) = \bar{C}(t)$, and (ii) a constraint or flux $\varphi(t)$,[4] where flux refers to some specific physical device, for example to a flow reactor. The first case is given, for example, by changes in the environment as there are periodic changes like day and night or seasons. In addition there are slow non-periodic changes or changes with very long periods like climatic change. Constraints and fluxes may correspond to unspecific or specific migration.[5] Considering time dependent carrying capacity and variable constraints simultaneously, we obtain

$$\frac{\mathrm{d}N_j}{\mathrm{dt}} = G_j(\boldsymbol{N}) - \frac{N_j}{K(t)}\,\varphi(t);\ j = 1, 2, \ldots, n\ . \tag{1.16}$$

Summation over the concentrations of all variants $\mathbf{X}_j$ and restricting the analysis to slowly changing environments – $K(t)$ varies on a time scale that is much longer than the time scale of population growth $C(t)$ – we can assume that the total concentration is quasi equilibrated, $C \approx \bar{C} = K$, and obtain a relation between the time dependencies of flux and total concentration:

$$\begin{aligned} \varphi(t) &= \sum_{i=1}^{n} G_j(\boldsymbol{N}) \ - \ \frac{\mathrm{d}C}{\mathrm{dt}} \quad \text{or} \\ C(t) &= C(0) + \int_0^t \left(\sum_{i=1}^{n} G_j(\boldsymbol{N}) - \varphi(\tau) \right) \mathrm{d}\tau\ . \end{aligned} \tag{1.17}$$

The proof for internal equilibration in growing populations is straightforward.

[4] There is a difference in the definitions of the fluxes ϕ and φ: $\phi(t) = \varphi(t)/C(t)$.

[5] *Unspecific migration* means that the numbers N_j of individuals for each variant $\mathbf{X}_j$ decrease (or increase) proportional to the numbers of individuals currently present in the population, $\mathrm{d}N_j = kN_j\,\mathrm{dt}$. *Specific migration* is anything else. In a flow reactor, for example, we have a dilution flux corresponding to unspecific emigration and an influx of one or a few molecular species corresponding to specific immigration into the reactor.

Theorem 1.2 (Equilibration in populations of variable size)**.** *Evolution in populations of changing size approaches the same internal equilibrium as evolution in populations of constant size provided the growth functions are homogeneous functions of degree γ in the variables N_j. Up to a transformation of the time axis, stationary and variable populations have identical trajectories provided the population size stays finite and does not vanish.*

Proof. Normalized variables, $x_i = N_i/C$ with $\sum_{i=1}^{n} x_i = 1$, are introduced in order to separate of population growth, $C(t)$, and population internal changes in the distribution of variants $\mathbf{X}_i$. From equations (1.16) and (1.17) with $C = \bar{C} = K$ and $N_j = Cx_j$ follows:

$$\frac{\mathrm{d}x_j}{\mathrm{dt}} = \frac{1}{C}\left(G_j\big(C\boldsymbol{x}\big) - x_j \sum_{i=1}^{n} G_i\big(C\boldsymbol{x}\big)\right);\ j = 1, 2, \ldots, n\ . \tag{1.18}$$

The growth functions are assumed to be homogeneous of degree γ in the variables[6] N_j: $G_j(\boldsymbol{N}) = G_j(C\boldsymbol{x}) = C^{\gamma}\, G_j(\boldsymbol{x})$. and we find

$$\frac{1}{C^{\gamma-1}}\frac{\mathrm{d}x_j}{\mathrm{dt}} = G_j(\boldsymbol{x}) \ - \ x_j \sum_{i=1}^{n} G_i(\boldsymbol{x});\ j = 1, 2, \ldots, n\ ,$$

which is identical to the selection equation in normalized variables for $C = 1$. For $\gamma = 1$ the concentration term vanishes and the dynamics in populations of constant and variable size are described by the same ODE. In case $\gamma \neq 1$ the two systems still have identical trajectories and equilibrium points up to a transformation of the time axis (for an example see section 4.2):

$$\mathrm{d}\tilde{t} = C^{\gamma-1}\ \mathrm{dt} \quad \text{and} \quad \tilde{t} = \tilde{t}_0 \ + \ \int_{\tilde{t}_0}^{\tilde{t}} C^{\gamma-1}(t)\ \mathrm{dt}\ ,$$

where $\tilde{t}_0$ is the time corresponding to $t = 0$ – commonly $\tilde{t}_0 = 0$. From equ. (1.18) we expect instabilities at $C = 0$ and $C = \infty$. □

[6] A homogenous function of degree γ is defined by $G(C\boldsymbol{x}) = C^{\gamma}G(\boldsymbol{x})$. The degree γ is determined by the mechanism of reproduction. For sexual reproduction according to Ronald Fisher's selection equation (2.9) we have $\gamma = 2$ [92]. Asexual reproduction discussed here fulfils $\gamma = 1$.

The instability at vanishing population size, $\lim C \to 0$, is, for example, also of practical importance for modeling drug action on viral replication. In the case of lethal mutagenesis [30, 31, 283] medication aims at eradication of the virus population, $C \to 0$, in order to terminate the infection of the host. At the instant of virus extinction equ. (1.9) is no longer applicable (see chapter 6.2).

1.4 Optimization

For the discussion of the interplay of selection and optimization we shall assume here that all fitness values f_j are different. The case of neutrality will be analyzed in chapter 10.3 and without loosing generality we rank them:

$$f_1 > f_2 > \ldots > f_{n-1} > f_n \; . \tag{1.19}$$

The variables $x_j(t)$ fulfil two time limits:

$$\lim_{t\to 0} x_j(t) = x_j(0) \; \forall \; j = 1, 2, \ldots, n \;\; \text{by definition, and}$$

$$\lim_{t\to\infty} x_j(t) = \begin{cases} 1 \text{ iff } j = 1 \\ 0 \; \forall \; j = 2, \ldots, n \; . \end{cases}$$

In the long time limit the population becomes homogeneous and contains only the fittest genotype $\mathbf{X}_1$. The process of selection is illustrated best by differential fitness, $f_j - \phi(t)$, the second factor in the ODE (1.14): The constraint $\phi(t) = \sum_{i=1}^{n} f_i x_i = \overline{f}$ represents the mean fitness of the population. The population variables x_l of all variants with a fitness below average, $f_l < \phi(t)$, decrease whereas the variables x_h with $f_h > \phi(t)$ increase. As a consequence the average fitness $\phi(t)$ is increasing too and more genotypes fall below the threshold for survival. The process continues until the fittest variant is selected. Since another view of optimization will be needed in chapter 2, we present another proof for the optimization of mean fitness without referring to differential fitness.

Theorem 1.3 (Optimization of mean fitness)**.** *The mean fitness* $\phi(t) = \bar{f} = \sum_{i=1}^{n} f_i x_i$ *with* $\sum_{i=1}^{n} x_i = 1$ *in a population as described by equ.* (1.14) *is non-decreasing.*

Proof. The time dependence of the mean fitness or flux ϕ is given by

$$\begin{aligned}\frac{\mathrm{d}\phi}{\mathrm{dt}} &= \sum_{i=1}^{n} f_i \frac{\mathrm{d}x_i}{\mathrm{dt}} = \sum_{i=1}^{n} f_i \left(f_i x_i - x_i \sum_{j=1}^{n} f_j x_j \right) = \\ &= \sum_{i=1}^{n} f_i^2 x_i - \sum_{i=1}^{n} f_i x_i \sum_{j=1}^{n} f_j x_j \; = \\ &= \overline{f^2} - \left(\overline{f}\right)^2 = \mathrm{var}\{f\} \geq 0 \; .\end{aligned} \tag{1.20}$$

Since a variance is always nonnegative, equ. (1.20) implies that $\phi(t)$ is a nondecreasing function of time. □

The condition $\mathrm{var}\{f\} = 0$ is met only by homogeneous populations. The one containing only the fittest variant $\mathbf{X}_1$ has the largest possible mean fitness: $\bar{f} = \phi_{\max} = f_1 = \max\{f_j; j = 1, 2, \ldots, n\}$. ϕ cannot increase any further and hence, it was been optimized by the selection process. The state of maximal fitness of population $\Upsilon = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$, $\boldsymbol{x}|_{\max\{\phi(\Upsilon)\}} = \{x_1 = 1, x_i = 0 \,\forall\, i = 2, \ldots, n\} = \boldsymbol{P}_1$, is the unique stable stationary state, and all trajectories starting from initial conditions with nonzero amounts of $\mathbf{X}_1$, $x_1 > 0$, have $\boldsymbol{P}_1$ as ω-limit. An illustration of the selection process with three variants and the trajectories are plotted on the unit simplex $\mathbb{S}_3^{(1)}$ is shown in figure 1.4.

Gradient systems [143, p.199] facilitate the analysis of the dynamics, they obey the equation

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{dt}} = -\mathrm{grad}\{V(\boldsymbol{x})\} = -\nabla V(\boldsymbol{x}) \tag{1.21}$$

and fulfil criteria that are relevant for optimization:

(i) The eigenvalues of the linearization of (1.21) evaluated at the equilibrium point are real.

(ii) If $\bar{\boldsymbol{x}}_0$ is an isolated minimum of V then $\bar{\boldsymbol{x}}_0$ is an asymptotically stable solution of (1.21).

(iii) If $\boldsymbol{x}(t)$ is a solution of (1.21) that is not an equilibrium point, then $V\big(\boldsymbol{x}(t)\big)$ is a strictly decreasing function and the trajectories are perpendicular to the constant level sets of V.

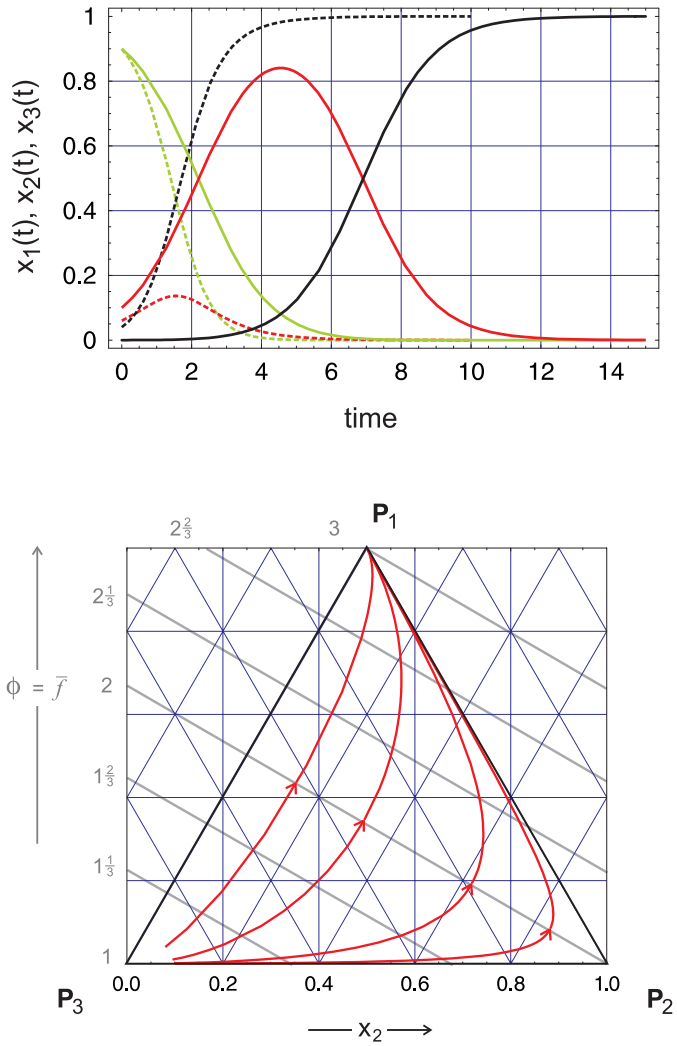


Figure 1.4: **Selection on the unit simplex.** In the upper part of the figure we show solution curves $\boldsymbol{x}(t)$ of equ. (1.15) with $n = 3$. The parameter values are: $f_1 = 3\,[\mathrm{t}^{-1}]$, $f_2 = 2\,[\mathrm{t}^{-1}]$, and $f_3 = 1\,[\mathrm{t}^{-1}]$, where $[\mathrm{t}^{-1}]$ is an arbitrary reciprocal time unit. The two sets of curves differ with respect to the initial conditions: (i) $\boldsymbol{x}(0) = (0.02, 0.08, 0.90)$, dotted curves, and (ii) $\boldsymbol{x}(0) = (0.0001, 0.0999, 0.9000)$, full curves. Color code: $x_1(t)$ black, $x_2(t)$ red, and $x_3(t)$ green. The lower part of the figure shows parametric plots $\boldsymbol{x}(t)$ on the unit simplex $\mathbb{S}_3^{(1)}$. Constant level sets of $\phi(\boldsymbol{x}) = \bar{f}$ are shown in grey. The trajectories refer to different initial conditions.

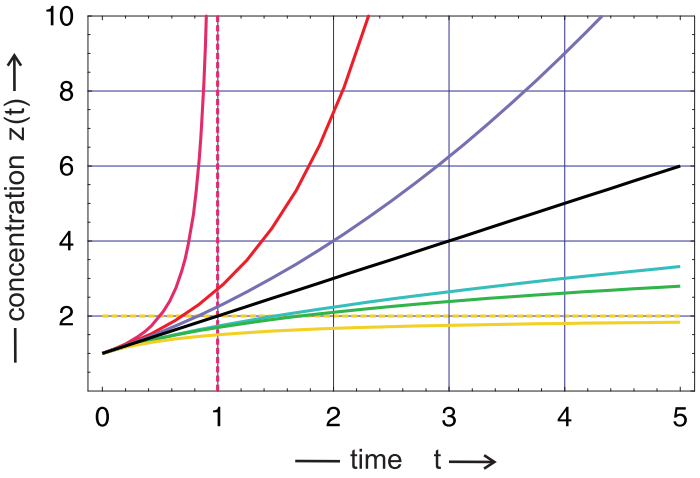


Figure 1.5: **Typical functions describing unlimited growth**. All functions are normalized in order to fulfil the conditions $z_0 = 1$ and $\mathrm{d}z/\,\mathrm{dt}|_{t=0} = 1$. The individual curves show hyperbolic growth ($z(t) = 1/(1-t)$; magenta; the dotted line indicates the position of the instability), exponential growth ($z(t) = \exp(t)$; red), parabolic growth ($z(t) = (1+t/2)^2$; blue), linear growth ($z(t) = 1+t$; black), sublinear growth ($z(t) = \sqrt{1+2t}$; turquoise), logarithmic growth ($z(t) = 1 + \log(1+t)$; green), and sublogarithmic growth ($z(t) = 1 + t/(1+t)$; yellow; the dotted line indicates the maximum value $z_{\max}$: $\lim_{t\to\infty} z(t) = z_{\max}$).

(iv) Neither periodic nor chaotic solutions of (1.21) do exist.

The relation between gradients systems and optimization is clearly seen from the first part of (iii): Replacing the minus signs in equ. (1.21) by plus signs reveals that $V\big(\boldsymbol{x}(t)\big)$ is non-decreasing and approaches a (at least local) maximum in the limit $t \to \infty$. As easily seen from figure 1.4 the trajectories of (1.14) are not perpendicular to the constant level sets of $\phi(\boldsymbol{x})$ and hence, equ. (1.14) is not a gradient system in the strict sense. With the definition of a generalized inner product corresponding to a Riemannian metric [261], however, the selection equation can be visualized as a generalized gradient and oscillations as well as deterministic chaos can be excluded [255].

1.5 Growth functions and selection

It is worth considering different classes of growth functions $z(t)$ and the behavior of long time solutions of the corresponding ODEs. An intimately related problem concerns population dynamics: What is the long time or equilibrium distribution of genotypes in a normalized population, $\lim_{t\to\infty} \boldsymbol{x}(t)$ provided the initial distribution has been $\boldsymbol{x}_0$? Is there a universal long time behavior, for example selection, coexistence or cooperation, that is characteristic for certain classes of growth functions?

Differential equations describing unlimited growth of the class

$$\frac{\mathrm{d}z}{\mathrm{dt}} = f \cdot z^n \tag{1.22}$$

will be compared here. Integration yields two types of general solutions for the initial value $z(0) = z_0$

$$z(t) = \left(z_0^{1-n} + (1-n)ft\right)^{1/(1-n)} \qquad \text{for } n \neq 1 \text{ and} \tag{1.22a}$$

$$z(t) = z_0 \cdot e^{ft} \qquad \text{for } n = 1\ . \tag{1.22b}$$

In order to make growth functions comparable we normalize them such that they fulfil the two conditions $z_0 = 1$ and $\mathrm{d}z/\,\mathrm{dt}|_{t=0} = 1$. For both equs. (1.22) this yields $z_0 = 1$ and $f = 1$. The different classes of growth functions, which are drawn in different colors in figure 1.5, are characterized by the following behavior:

(i) Hyperbolic growth requires $n > 1$; for $n = 2$ it yields a solution curve of the form $z(t) = 1/(1-t)$. Characteristic is the existence of an instability in the sense that $z(t)$ approaches infinity at some critical time, $\lim_{t\to t_{\mathrm{cr}}} z(t) = \infty$ with $t_{\mathrm{cr}} = 1$. The selection behavior of hyperbolic growth is illustrated by the Schlögl model:[7] $\mathrm{d}z_j/\,\mathrm{dt} = f_j z_j^2$; $j = 1, 2, \ldots, n$. Depending on the initial conditions each of the replicators $\mathbf{X}_j$ can be selected. $\mathbf{X}_m$ the species with the highest replication

[7] The Schlögl model is tantamount to Fisher's selection equation with diagonal terms only: $f_j = a_{jj}$; $j = 1, 2, \ldots, n$ [242].

parameter, $f_m = \max\{f_i; i = 1, 2, \ldots, n\}$ has the largest basin of attraction and the highest probability to be selected. After selection has occurred a new species $\mathbf{X}_k$ is extremely unlikely to replace the current species $\mathbf{X}_m$ even if its replication parameter is substantially higher, $f_k \gg f_m$. This phenomenon is called *once-for-ever selection*.

(ii) Exponential growth is observed for $n = 1$ and described by the solution $z(t) = e^t$. It represents the most common growth function in biology. The species $\mathbf{X}_m$ having the highest replication parameter, $f_m = \max\{f_i; i = 1, 2, \ldots, N\}$, is always selected, $\lim_{t\to\infty} z_m = 1$. Injection of a new species $\mathbf{X}_k$ with a still higher replication parameter, $f_k > f_m$, leads to selection of the fitter variant $\mathbf{X}_k$ (fig.1.3).

(iii) Parabolic growth occurs for $0 < n < 1$ and for $n = 1/2$ has the solution curve $z(t) = (1 - t/2)^2$. It is observed, for example, in enzyme free replication of oligonucleotides that form a stable duplex, i.e. a complex of one plus and one minus strand [295]. Depending on parameters and concentrations coexistence or selection may occur [311].

(iv) Linear growth follows from $n = 0$ and takes on the form $z(t) = 1 + t$. Linear growth is observed, for example, in replicase catalyzed replication of RNA at enzyme saturation [17].

(v) Sublinear growth occurs for $n < 0$. In particular, for $n = -1$ gives rise to the solution $z(t) = (1 + 2t)^{1/2} = \sqrt{1 + 2t}$.

In addition we mention also two additional forms of weak growth that do not follow from equ. (1.22):

(vi) Logarithmic growth that can be expressed by the function $z(t) = z_0 + \ln(1 + ft)$ or $z(1) = 1 + \ln(1 + t)$ after normalization, and

(vii) sublogarithmic growth modeled by the function $z(t) = z_0 + ft/(1 + ft)$ or $z(t) = 1 + t/(1 + t)$ in normalized form.

Hyperbolic growth, parabolic growth, and sublinear growth constitute families of solution curves that are defined by a certain parameter range (see

figure 1.5), for example a range of exponents, $n_{\text{low}} < n < n_{\text{high}}$, whereas exponential growth, linear growth and logarithmic growth are critical curves separating zones of characteristic growth behavior: Logarithmic growth separates growth functions approaching infinity in the limit $t \to \infty$, $\lim_{t\to\infty} z(t) = \infty$ from those that remain finite, $\lim_{t\to\infty} z(t) = z_\infty < \infty$, linear growth separates concave from convex growth functions, and exponential growth eventually separates growth functions that reach infinity at finite times from those that don't.

2. Mendel's genetics and recombination

Darwin's principle of natural selection was derived from a wealth of observed adaptations that he had made during all his life and in particular on a journey all around the world, which he made as the naturalist on H.M.S. Beagle. Although adaptations are readily recognizable in nature with educated eyes, very little was evident about the mechanisms of inheritance except perhaps the general principle that children resemble their parents to some degree. Similarity in habitus manifests itself nowhere so clearly as with identical twins, and this was, of course, already noticed long time before genetics has been discovered and analyzed. Although twins were of interest to scholars since the beginnings of civilization, for example in the fifth century B.C. the Greek physician Hippocrates had been studying similarity in the course of diseases in twins, the modern history of twin research was initiated only since the nineteenth century by the polymath Sir Francis Galton, who was a cousin of Charles Darwin. The lack of insight into the mechanisms of inheritance, however, caused him and many other scientists and physicians afterwards – among them also the population geneticist Ronald Fisher [91] – to miss the difference between monozygotic (MZ) or identical and dizygotic (DZ) or fraternal twins. Before Fisher's failure, however, this difference had been recognized already by the German physician and pioneer of population genetics Wilhelm Weinberg [303] and later rediscovered and documented by the German physician Hermann Werner Siemens [263].

Darwin's ideas on inheritance focussed on the concept of *pangenesis*, which assumed that tiny particles from cells, so called *gemmules*, are transmitted from parents to offspring and maternal and paternal features are blended in the progeny. Pangenesis, however, was wrong in two important aspects: (i) Not all cells contribute to inheritance only the germ cells [304] and (ii) inheritance occurs in discrete packages nowadays called genes, many features, for example the colors or leaves, flowers or fruits, are discrete rather

than continuously varying. Here we shall start with a discussion of Gregor Mendel's experiments on *Pisum sativum*, the garden pea [206], and *Hieracium*, the hawkweed [207], and after that introduce elementary population genetics and, in particular, we derive the Hardy-Weinberg equilibrium, and analyze Fisher's selection equation and the fundamental theorem. Finally we shall discuss Fisher's criticism on Mendel's work.

2.1 Mendel's experiments

The Augustinian friar Gregor Mendel performed a series of experiments with plants under controlled fertilization (For a detailed outline of Mendel's experiments and patterns of inheritance in general see [125, pp.27-66]. Luckily Mendel was choosing the garden pea, Pisum sativum as the object of his studies. His works are remarkable for at least two reasons: (i) Mendel improved the experimental pollination technique in such a way that unintended fertilization could be excluded (Among more than 10 000 plants, which were carefully examined, only in a very few cases an indubitable false impregnation by foreign pollen had occurred), and (ii) he discovered a statistical law and therefore he had to carry out a sufficiently large number of individual experiments before the regularities became evident.

Mendel's contributions to evolutionary biology were twofold:
(i) He discovered two laws of inheritance, Mendel's first law called the law of segregation – the hereditary material is cut into pieces that represent individual characters in the offspring, and Mendel's second law called independent assortment – the hereditary characters from father and mother come into a pool are combined anew without reference to their parental combinations.
(ii) By careful planning and recording of experiments he found two modes of hereditary transmission: *Dominance* – one of the two parental features is reproducibly transmitted to the offspring whereas the second one disappears completely in the first generation (F1) – and recessiveness – a feature that has disappeared in the first generation will show up again if hybrid individuals of the first generation are crossed among each other (F2).

Gregor Mendel was choosing seven characters for experimental recording:

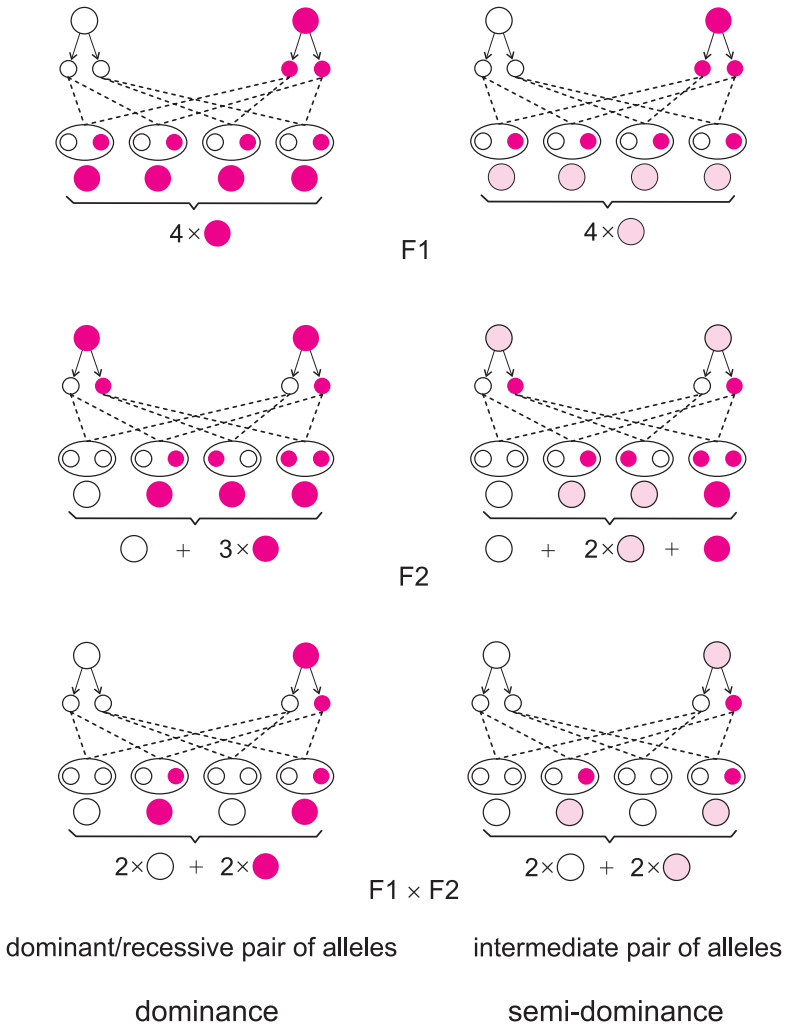


Figure 2.1: **Mendelian genetics**. The rules of genetic inheritance are illustrated by means of a simple sketch. Flowers appear in two colors, white and red. Two plants that are homozygous at the color locus are cross-fertilized and yield a generation of heterozygotes (F1). Cross-fertilization of F1 plants yields the second generation F2. Two cases are distinguished: dominance (lhs) and semi-dominance (rhs). In case of dominance the heterozygote exhibits the same features as the homozygote of the dominant allele (red color in the example), and this leads to a ratio of 1:3 in the phenotypes of the second generation F2. The heterozygote of an intermediate pair of alleles shows an intermediate feature (pink color) and then the ratio of phenotypes is 1:2:1.

(i) *The difference in the form of ripe seeds.* Round or roundish versus irregularly angular and deeply wrinkled, studied in 60 fertilizations on 15 plants.

(ii) *The difference in the color of the seed endosperm.* Pale yellow, bright yellow or orange versus more or less intense green, studied in 58 fertilizations on 10 plants.

(iii) *The difference in the color of the seed-coat.* White or gray, gray-brown or leather-brown with or without violet spotting, studied in 35 fertilizations on 10 plants.

(iv) *The difference in the form of the ripe pods.* Simply inflated versus deeply constricted and more or less wrinkled, studied in 40 fertilizations on 10 plants.

(v) *The difference in the color of the unripe pods.* Light to dark green versus vividly yellow, studied in 23 fertilizations on 5 plants.

(vi) *The difference in the position of the flowers.* Axial (distributed along the main stem) versus terminal (bunched at the top of the stem), studied in 34 fertilizations on 10 plants.

(vii) *The difference in the length of the stem.* Long (1.8 to 2.1 m) versus short (25 to 50 cm) distinguishable for healthy plants grown in the same soil, studied in 37 fertilizations on 10 plants.

Mendel first created hybrids from plants with opposite forms of the seven characters and these hybrids constitute the generation F1, which is genetically homogeneous. Crossings of two (genetically identical) individuals of generation F1 leads three different genotypes in the F2 generation. For all the characters he has been studying he observed two different phenotypes with a ratio around 3:1 (table 2.1). Mendel's correct interpretation is illustrated in fig. 2.1: All (diploid) organisms carry two alleles at every locus, they are homozygous if the two alleles are identical and heterozygous if the alleles are different. Cross-fertilization of two homozygous plants yields identical

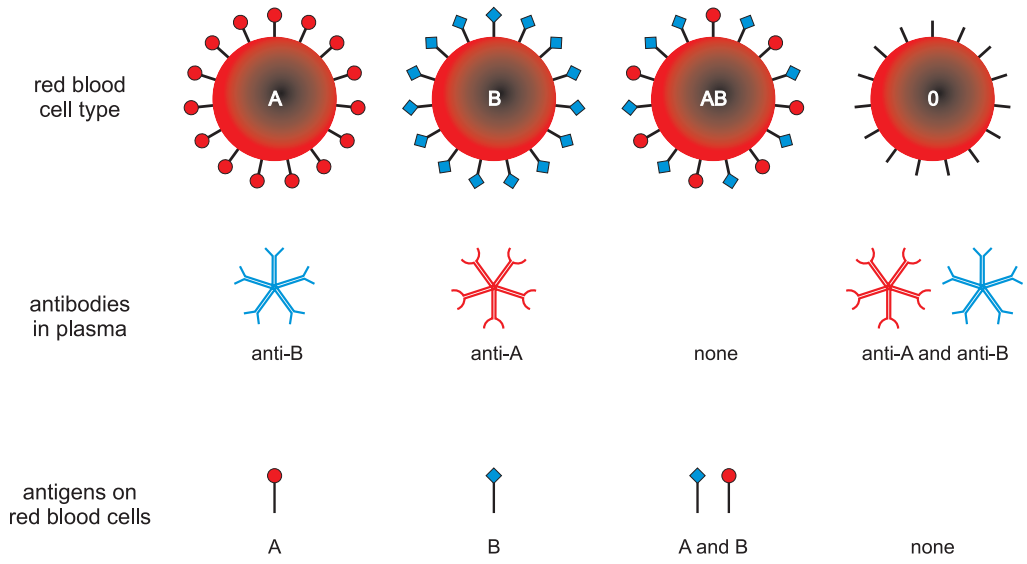


Figure 2.2: **Human blood types as an example for codominance**. The sketch shows the *erythrocytes* of the four human blood types with the antigens expressed on the cell surface (top row). In the middle row we see the antibodies that are present in the blood plasma after contact with the corresponding antigens. No antibodies are developed against antigens that are recognized as *self* by the immune system (bottom row).

offspring. Since this is not the case if one or both parents are heterozygous, this criterion can be used to identify homozygous individuals. When two identical genotypes of the F1 generation are cross-fertilized, three different genotypes are obtained, the two homozygotes and two heterozygotes.[1]

Mendel's observation implied that the heterozygotes and one of the two homozygotes developed the same phenotype. All seven characters correspond to this situation and the following forms were present a higher frequency:

(i) the round or roundish form of the seeds,

[1]In Mendelian genetics the two heterozygotes are indistinguishable because it is assumed that the same phenotype is formed irrespectively whether a particular allele of an autosome come from the father or from the mother. All chromosomes except the sex chromosomes are autosomes and they are present in two copies in a diploid organism.

(ii) the yellow color of the endosperm,

(iii) the gray, gray-brown or leather-brown color of the seed coat,

(iv) the simply inflated form of the ripe pods,

(v) the green coloring of the unripe pod,

(vi) the axial distribution of the flowers along the stem, and

(vii) the long stems.

The figure shows in addition the ratios of phenotypes when one individual of the F1 generation is cross fertilized with a homozygous plant of the F2 generation. Later such an allele pair has been denoted as dominant-recessive. In table 2.1 we show the detailed results of Mendel's experiments and point a two features that a typical for a statistical law: (i) the large number of repetitions, which are necessary to be able to recognize the regularities and (ii) the rather small deviations from the ideal ratio three. In section 2.6 we shall analyze Mendel's data by means of the χ^2-test, a statistical reliability test that has been introduced around nineteen hundred by the Karl Pearson.

From Mendel's experiment we conclude that every diploid organism carries two copies of each (autosomal) gene. The copies are separated during sexual reproduction and combined anew. Alleles shall be denoted by sans-serif letters, in a dominant-recessive allele pair we shall denote the dominant allele by an upper-case letter and the recessive allele by a lower-case letter, A and a, respectively. The four zygote are then AA, Aa, aA, and aa where the first three genotypes express the same phenotype. Although dominance is by far the more common feature in nature, other form exist and they are also familiar to careful observers and naturalists.

Incomplete dominance or *semi-dominance* is a form of intermediate inheritance in which one allele for a specific trait is not completely dominant over the other allele, and a combined phenotype is the results (fig. 2.1, rhs): The phenotype expressed by the heterozygous genotype is an intermediate of the phenotypes of the homozygous genotypes. For example, the color of the

Table 2.1: **Results of Gregor Mendel's experiments with the garden pea (*pisum sativum).*** The list contains all results of Mendel's crossing experiments in which the parents different in one character. the ration between phenotypes is very close to three that is the ideal ration derived from Mendel's principle of inheritance.

Char.	Parental phenotype	F1	F2	F2 ratio
1	round × wrinkled seeds	all round	5174 / 1859	2.96
2	yellow × green seeds	all yellow	6022 / 2001	3.01
3	purple × white petals	all purple	705 / 244	3.15
4	inflated × pinched pods	all inflated	882 / 299	2.95
5	green × yellow pods	all green	428 / 152	2.82
6	axial × terminal flowers	all axial	651 / 207	3.14
7	long × short stems	all axial	787 / 277	2.84

snapdragon flower in homozygous plants is either red or white. When the red homozygous flower is cross-fertilized with the white homozygous flower, the result yields a pink snapdragon flower. A similar form of incomplete dominance is found in the four o'clock plant where in pink color is produced when true bred parents of white and red flowers are crossed. The lack of dominance is expressed in the notation through choosing upper-case letters for both alleles, for example the alleles A and B give rise to the four genotypes AA, AB, BA, and BB whereby the two heterozygotes produce the same phenotype. When plants of F1 generation are self pollinated the phenotypic and genotypic ratio of the F2 generation will be same and is 1:2:1 (Red:Pink:White), because three phenotypes can be distinguished. The intermediate color commonly is a result of pigment concentration: One allele, R, produces the red color, the other one 0 does not give rise to color expression, and then RR has twice as much red pigment than R0.

Codominance is another genetic mechanism that leads to two alleles on an equal footing. The allelic products coexist in the phenotype and the contributions of both alleles at the single locus are clearly visible and do not

overpower each other. Codominance is different from incomplete or semi-dominance, where the quantitative interaction of allele products produces an intermediate phenotype like the pink snapdragon obtained by crossing homozygous plants with red and white flowers. In case of codominance the hybrid genotype derived from a red and a white homozygous flower will produce offspring that have red and white spots. As an well studied example of codominance we mention the human AB0 blood type system, because it has a very simple explanation on the molecular level (fig. 2.2). Three alleles from six diploid genotypes which develop four phenotypes:

genotype	phenotype
AA	A
BB	B
00	0
AB	AB
A0	A
B0	B

Codominance of the two alleles A and B leads to the blood type AB where the two alleles coexist in the phenotype. The explanation is straightforward: The red blood cells called *erythrocytes* express characteristic antigens on the cell surface and antibodies are developed against non-self antigens (fig. 2.2). The blood types determine a possible antigen-antibody reaction that causes mixed blood samples to *agglutinate* or forms blood clumps. If this happens after a blood transfusion the patient develops a very serious usually lethal acute hemolytic reaction. Red blood cell compatibilities are readily derived from fig. 2.2: AB type individuals can receive blood from any group but can donate only to other AB individuals, 0 type individuals in contrary can donate blood to all blood types but receive blood only from individuals of the 0 group, A group individuals can receive blood from A and 0 type individuals, and analogously B blood is compatible with samples from B and 0 type individuals.

As an example for dominance in human genetics we consider the rhesus (Rh) blood group system. It is highly complex and dealing with about fifty

antigens, out of which five, D, C, c, E, and e being the most important ones. The term rhesus (Rh) factor is commonly used for the D/d allele pair on one locus. Rh positive and Rh negative refer to the D antigen only, because the d allele expresses no antigen (like the 0 allele):

genotype	phenotype
DD	Rh D positive
Dd	Rh D positive
dd	Rh D negative

The rhesus factor plays a role in blood transfusion but is also responsible for *Rh D hemolytic disease of the newborn*. If the genotype of the mother is dd (Rh negative) sensitization to Rh D antigens caused by feto-maternal blood transfusion through the placenta can lead to the production of maternal anti-D antibodies that will effect any subsequent pregnancy and lead to the disease in case the baby is Rh D positive. The vast majority of cases of Rh disease can be prevented by modern antenatal care through injections of anti-D antibodies called *Rho(D) immune globulin*. The prevalence of Rh negative people varies substantially in different ethnic groups. The Rh negative phenotype is most common ($\approx 30\,\%$) among the Basque people and quite common among the other Europeans ($\approx 16\,\%$) and very rare ($\approx 1\,\%$ and less) in Asian and Native American populations. African Americans are intermediate ($\approx 7\,\%$).

2.2 The mechanism of recombination

Recombination of packages of genetic information during sexual reproduction was kind of a mystery as long as the mechanism at the molecular level was unknown or unclear. Cell biology and in particular the spectacular development of molecular biology shed light on the somewhat obscure seeming partitioning of genetic information into packages. Already August Weismann had the correct idea that there is a fundamental difference between cells in the germ-line and somatic cells [304, 305] and inheritance is based on germ-line cells alone. The germ-line cells fall into two classes, sperms and eggs,

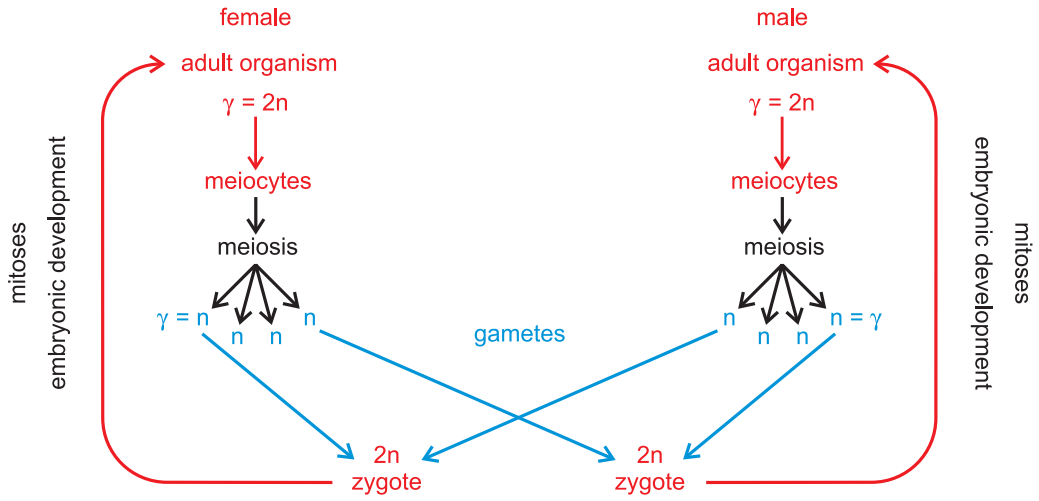


Figure 2.3: **The life cycle of a diploid organism**. The life cycle of a typical diploid eukaryote consists of a haploid phase with $\gamma = n$ chromosomes (blue) that is initiated by meiosis and ends with the fusion of a sperm and an egg cell in order to form a diploid zygote. During the rest of the life cycle each somatic cell of the organism has $\gamma = 2n$ chromosomes in $n - 1$ autosomal pairs and the sex chromosomes (red). Special cell lines differentiate into meiocytes, which undergo meiosis and form the gametes.

which differ largely in the amount of cytoplasm that they contain: In the sexual union of sperm and egg forming a zygote the egg contributes almost the entire cytoplasm. The nuclei of egg and sperm cells are of approximately the same size and therefore the nuclei were considered as candidates for harboring the structures that are responsible for inheritance. A sketch of the typical diploid life cycle with a long diploid phase and a short haploid stage providing the frame for sexual reproduction is shown in Fig. 2.3. In the eighteen eighties the German biologist Theodor Boveri demonstrated that within the nucleus the chromosomes were the vectors of heredity. It was also Boveri who pointed out that Mendel's rules of inheritance are consistent with the observed behavior of chromosomes and developed independently from Walter Sutton in 1902 the chromosome theory of inheritance. The ultimate proof of the role of chromosome was provided by the American geneticist Thomas

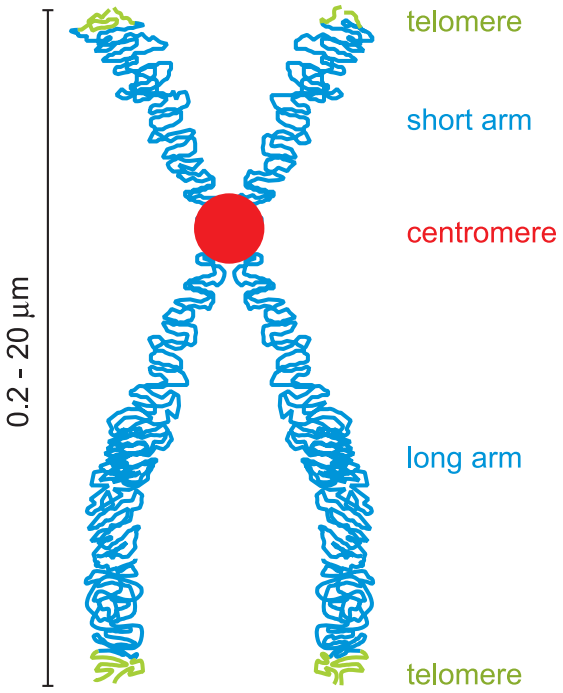


Figure 2.4: **Sketch of a replicated and condensed eukaryotic chromosome**. Shown are the two sister chromatids in the meta phase after the synthetic phase. The centromere (red) is the point where the two chromatids touch and where the microtubules attach. The ends of the chromatids (green) are called telomeres and carry repetitive sequences that protect the chromosomes against damage.

Hunt Morgan who started with systematic crossing experiments with the fruit fly *drosophila* around 1910.

2.2.1 Chromosomes

Chromosomes are complex structures consisting of DNA molecules and proteins. The necessity to organize DNA structure becomes evident from considering its size: The DNA molecule of a human consists of $2\times3\times10^9$ base pairs and in fully stretched state the double-helical DNA molecule would be $6\times10^9 \cdot 0.34\,\mathrm{nm} \approx 2\,\mathrm{m}$ long. Clearly, such a long molecule can only be processed successfully in an compartment with the diameter of a eucaryotic cell when properly condensed to smaller size. Mammalian cells vary considerably in size: Among the smallest are the red blood cells with a diameter of about $0.76\,\mu$m and among the largest cells are are the nerve cells that span a giraffe's neck and which can be longer than 3 m. Analogously human nerve cells may be as long as 1 m. The size of the average human cell, however, lies

in the range $1 \leq \mathrm{d} \leq 10\,\mu$m. In Fig. 2.2 we present a sketch of a duplicated chromosome after completed condensation. The condensation of the DNA leads to a length contraction by six orders of magnitude, i.e. from 2 m to about $2\,\mu$m. It occurs in several steps and involves histones and other proteins. Histones are positively charged (basic) proteins that bind strongly to the negative charged DNA molecule. Their sequence and structure is highly conserved in evolution. Here we give an idea only of the first steps, which is the formation of *chromatin* from core histones, linker histones, and DNA. An protein octamer built from two molecules each of the histones H2A, H2B, H3, and H4 forms the core of a nucleosome around which the DNA is wrapped twice. The resulting structure looks like beads on a string an has a diameter of $d \approx 10$ nm. Linker histones arrange nucleosomes forming a solenoid structure with six nucleosomes symmetrically arranged in one complete turn and a diameter of $d \approx 10$ nm. There are three homologous linker histones, H1, H5 and H1°, which apparently can replace each other and have additional specific functions [278]. With the help of scaffold proteins the solenoid structure is condensed further during *interphase* (see Fig. 2.5) yielding the *active chromosome* that is again condensed further through addition of more scaffold proteins, and eventually the *metaphase* chromosome is formed that is ready for cell division.

The numbers of chromosomes is variable and characteristic for species. They are divided into autosomes and sex chromosomes (subsection 2.2.2 and vary substantially in size. Human cells have 46 chromosomes, 22 pairs of autosomes and on pair of sex chromosomes. Chromosome 1 is the largest, it is almost 250 million base pairs long and carries to present-day knowledge 4 316 genes. Chromosome 21 is the smallest human chromosome, it is 47 million bases long and codes for 300 to 400 genes. Comparison of chromosome numbers in different species shows. Our closest relatives, Gorillas and chimpanzees have 48 chromosomes, domestic cats 38, dogs 78, cows 60, and horses 64. The variation among fishes is remarkable: Fugu has the smallest genome – only 392.4 million base pairs, which is about 1/8 of the human genome – and 44 chromosomes, guppy the popular aquarium fish 46 and the goldfish 100-104. Somewhat more complex is the chromosome number with birds:

The domestic pigeon has 18 large chromosomes and 60 microchromosomes and similarly chicken with 10 chromosomes and 60 microchromosomes, and eventually the pretty bird kingfisher 132 total. The chromosome numbers in plants are highly variable as well: the thale cress, *arabidopsis thaliana*, has 10 chromosomes and the pea, *pisum sativum*, 24, and the pineapple 50. The well-studied yeast, *saccharomyces cerevisiae*, has 32 chromosomes. Finally, we make a glance on prokaryotes. The eubacterium Escherichia coli has a single circular chromosome, which when fully stretched is many orders of magnitude larger than the cell itself, but no histones like other eubacteria. DNA is condensed mainly by supercoiling and the process is assisted by specific enzymes called topoisomerases [162]. Topoisomerases, in general, resolve the topological problems associated with DNA replication, transcription, recombination, and chromatin remodeling in a trivial but highly efficient way: They introduce temporary single- or double-strand breaks into DNA, unwind, and ligate again.

After DNA replication and condensation into chromosomes the two *sister chromatids* have a long and a short arm (Fig. 2.2) and they are joined at the *centromere* that is also the point of attachment of the *microtubules*, which organize chromosome transport during cell division. In order to prevent loss of DNA ends during cell division at the tips, the chromosomes carry *telomeres*, which are stretches of short repeats of oligonucleotides that can be understood as disposable buffers blocking the ends of the chromosomes. In case of vertebrates the repeat in the telomeres is **TTAGGG**. Part of the telomeres are consumed during cell division and replenished by an enzyme called *telomerase reverse transcriptase*. Cells, which have completely consumed their telomeres, are destroyed by apoptosis and in rare cases they find ways of evading programmed destruction and become *immortal* cancer cells. In 2009 Elizabeth Blackburn, Carol Greider, and Jack Szostak were awarded the Nobel Prize in Physiology and Medicine for the discovery of how chromosomes are protected by telomeres and the enzyme telomerase. The telomeres are tightly bound to the inner surface of the nuclear envelope during prophase 1 (see Fig. 2.6) and play an important role in pairing homologous chromosomes during meiosis.

2.2.2 Chromosomes and sex determination

A diploid organism (Fig. 2.3) carries $\gamma = 2n$ chromosomes forming $n-1$ pairs of autosomes and the pair of sex (determining) chromosomes. Sex is basic to diploid life and therefore the fact that there are several entirely different sex-determination systems has a kind of strange appeal. Three different chromosomal systems are known, one system changes the common haploid-diploid relation, and others invoke external parameters:

system	female	male	species
XX/XY	XX	XY	almost all mammals including man,
			some insects (*drosophila*), some plants, ...
XX/X0	XX	X0	insects, ...
ZW/ZZ	ZW	ZZ	birds, some fish, some reptiles and insects, ...
haplo-	$2n$	n	hymenoptera (most), ...
diploid			spider mites, bark beetles, rotifers, ...
temper-	warm	cold	some reptiles, few birds , ...
ature	medium	extreme	other reptiles, ...
infection	infected	not	butterflies (*Wolbachia* infection), ...
		infected	some nematodes, ...

Since the XX/XY sex-determination found in man is almost universal among mammals one is inclined to consider it as the only sex-determining system in nature but this is utterly untrue. The XX/XY-determination is wide spread in nature and this helps to believe it is the only one. The XX/X0-determination can be visualized as a XX/XY-system in which the already smaller Y-chromosome has been ultimately lost. An intermediate situation is found with the fruit-fly drosophila: In some variants (or species) the male carries a Y-chromosome whereas it has none in other variants. In the ZW/ZZ sex-determination system the female rather than the male carries the two different sex chromosomes. In the haplodiploid sex-determination system the male is haploid and the entire kinship relations are different from those in the conventional diplodiploid systems. The coefficients of relationship for parent and offspring expressed in the percentage of shared genes ($1 \equiv 100\,\%$) are [317]:

	haplodiploid		diplodiploid	
relative	female	male	female	male
daughter	1/2	1	1/2	1/2
son	1/2	–	1/2	1/2
mother	1/2	1/2	1/2	1/2
father	1/2	–	1/2	1/2
identical twin	–	–	1	1
full sister	3/4	1/2	1/2	1/2
full brother	1/4	1/2	1/2	1/2

In the case of haplodiploidy sisters are more closely related than in diploid-diploid organisms and this has been used as support for the frequent occurrence of eusociality in hymenoptera, in particular bees, wasps, and ants [131, 132]. Kinship in haplodiploid organisms as explanation for colony formation and altruistic behavior had one major problem: Termites are diploid organisms and form gigantic colonies with a complex caste system. Recently it was shown that the evolution of eusociality can be explained and modeled mathematically by means of natural selection [225]. Sex determination by nest temperature

2.2.3 Mitosis and meiosis

Here we can present a few some basic facts of this very extensive field of cytology and molecular biology, which is outlined in more detail in text books (for example, [5] and [125]) and which represents also a discipline of cutting-edge research [222]. From Fig. 2.3 follows that a diploid organism needs – at least – two types of cell divisions: (i) A division mechanism, which in general creates two identical diploid cells from one diploid precursor cell, and (ii) a mechanism, which creates haploid cells from a diploid precursor in order to allow for the formation of a diploid zygote through merging of two haploid cells during mating. The two most common and almost universal natural cell division mechanisms, mitosis and meiosis, are sketched in Figs. 2.5 and 2.6. In general, both mechanisms are symmetric in the sense that the two or four daughter cells are equivalent. Asymmetric cell divisions in the sense that the

offspring cells are intrinsically different play a role in embryonic development [139], in particular with stem cells [177]. As an example we mention the nematode *Caenorhabditis elegans* where several successive asymmetric cell divisions in the early embryo are critical for setting up the anterior/posterior, dorsal ventral, and left/right axes of the body plan [119].

In *mitosis* one diploid cell divides into two diploid cells that – provided no accident has happened – carry the same genetic information as the parental cell and the over-all process is simple duplication (Fig. 2.5). Homologous chromosomes behave independently during mitosis – and this distinguishes it from meiosis. The problem that has to be solved, nevertheless, is of formidable complexity: A molecule of about 2 m length has to be copied in a cell of about 10 μm diameter and than divided equally during cell division. Long before the advent of molecular biology cell division has been studied extensively by means of light microscopy and the different stages shown in Fig. 2.5 were distinguished. DNA replication takes place in the interphase nucleus and each chromosome is transformed into two identical *sister chromatids*, in *prophase* we have the sister chromatids in perfect alignment and ready for cell division then the nuclear membrane dissolves and in *metaphase* the chromosomes migrate to the equator of the cell. Microtubules form and attach to the centromeres (Fig. 2.2) in *anaphase* and *telophase* the sister chromatids are pulled apart and to the two opposite poles of the cell. At the end of telophase the cell splits into two daughter cells and nuclear membrane are formed in both cells.

Meiosis is initiated like mitosis by DNA replication but instead of cell division two organized cell divisions follow with no second replication phase in between. Accordingly one diploid cell is split into four haploid gametes during meiosis. The major difference between the two cell division scenarios occurs in prophase 1 and metaphase 1: The two duplicated homologous chromosomes are pairing and crossover between the four chromatids is disentangled by *homologous recombination*.[2] During prophase 1 the tight binding

[2]Homologous recombination is the precise notion of recombination during meiosis since there are also other forms of recombination in the sense of exchange of genetic material, for example, with bacterial conjugation or multiple virus infection of a single cell.

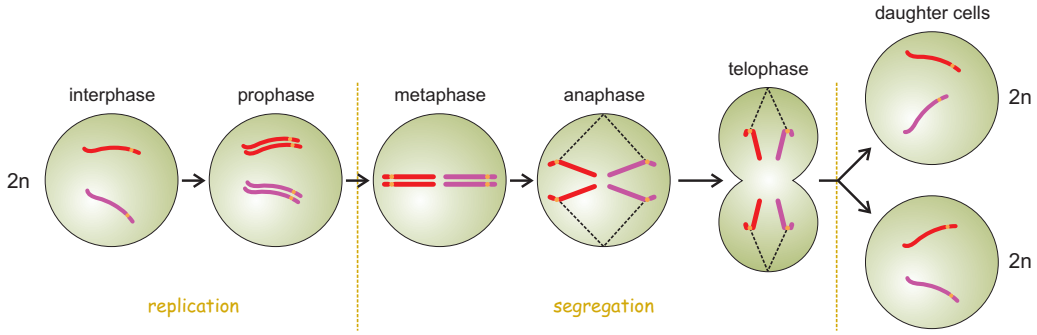


Figure 2.5: **Mitosis**. Mitosis is the mechanism for the division of nuclei associated with the division of somatic cells in eukaryotes. During interphase that comprises of three stages of the cell cycle – gap 1 (G1), synthesis (S), and gap2 (G2) – the DNA of each chromosome replicates and each chromosome is transformed into two sister chromatids, which lie side by side. The mitosis stage of the cell cycle (M) starts with *prophase* when the sister chromatids become visible in the light microscope. During *metaphase* the sister chromatid pairs are moving to the equatorial plane of the cell. In *anaphase* microtubules being part of the *nuclear spindle* attach to the centromeres (small orange balls in the sketch), separate the sister chromatids, and pull them in opposite direction towards the cellular poles. In *telophase* the separation is completed and a nuclear membrane forms around each nucleus and cell division completes mitosis. The sketch shows the fate of a single chromosome, which is present in two differently marked copies (red and bright violet) that appear in identical form in the two daughter cells. Mitosis produces two (in essence) genetically identical diploid (2n) daughter cells from one diploid cell (2n). Here and in Fig. 2.6 we do not show the nuclear membrane in order to avoid confusion. During interphase, the first part of prophase, and in the daughter cells the compartment shown is the nucleus whereas the circles mean the entire cell during the stages of segregation.

between the sister chromatids is weakened and eventually resolved through the formation of a large complex in which all four chromatids of the two duplicated homologous chromosomes are aligned by means of an extensive protein machinery. This process is slow as prophase 1 may occupy 90 % or more

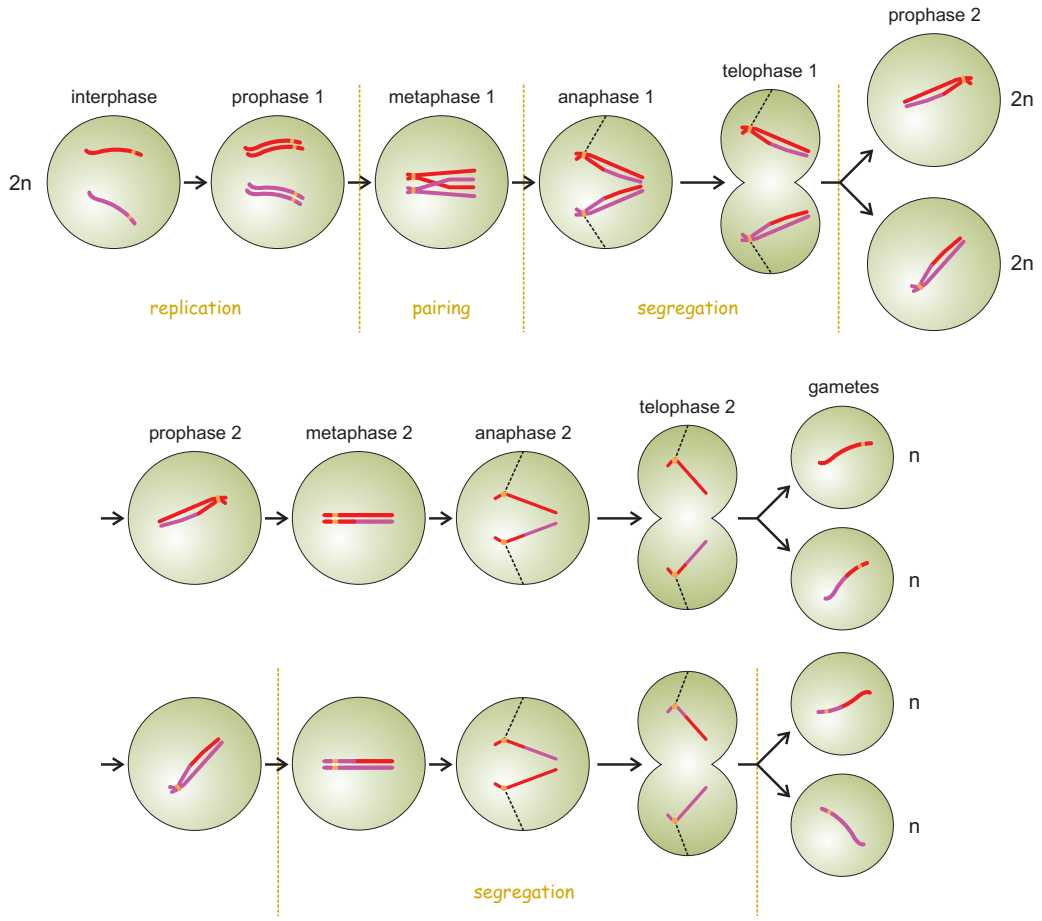


Figure 2.6: **Meiosis**. Meiosis in the mechanism by which diploid organisms produce haploid germ cells from a diploid precursor cell. The process is initiated like mitosis in *prophase 1* but then in *metaphase 1* the duplicated chromosomes are paired yielding a four chromatid complex, which is the stage where homologous recombination occurs (see Fig. 2.7). We show one crossing over of DNA double strands that is disentangled by recombination. Then follow two divisions without a synthetic phase – *anaphase 1*→*telophase 1*→*prophase 2*→*metaphase 2*→*anaphase 2*→*telophase 2* – and eventually after the second division we end up with four different gametes. The sketch shows the fate of a chromosome pair, which is initially present in two differently marked homologous copies (red and bright violet) during one DNA replication and two consecutive divisions into four daughter cells. Meiosis produces four genetically different haploid gametes (n) from one diploid cell (2n).

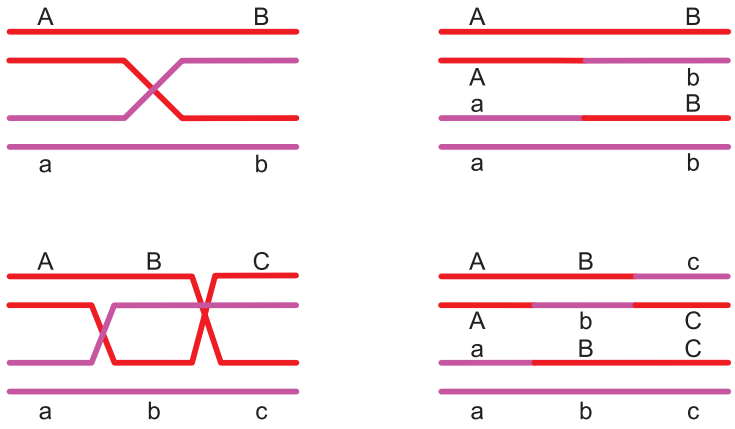


Figure 2.7: **Crossover and recombination**. Crossover occurs in the meiotic prophase 1 or metaphase 1 during pairing of homologous stretches from all four chromatids. Both forms single crossover and multiple crossover shown in the upper and lower parts of the figure, respectively, are possible and in general all four strands may be involved in it. Resolution of crossover through special mechanisms involving breaking and linking of the DNA double strands in chromatids leads two recombination shown on the rhs of the figure. Since at least one crossover is obligatory in meiosis – if no crossover occurs the process is arrested in metaphase 1, all four haploid gametes are genetically mixed and different unless the diploid organism has been homozygous in all genes.

of the total time of meiosis. The tightly aligned homologs form crossovers that can be seen as chiasmata in metaphase 1. Crossovers are resolved leading to recombination in the four chromatids and four different chromosomes are formed (For the sake of simplicity only one crossover event is shown in Fig. 2.6). In anaphase 1 and telophase 1 the eventually recombined duplicated homologous chromosomes segregate and enter after cell division as two separate diploid cells into the second division cycle initiated through prophase 2. Metaphase 2 is analogous to the metaphase in mitosis: homologous chromosomes align in the region of the cell equator segregate during anaphase 2 and telophase 2 and finally end up in four genetically mixed haploid gametes. Incase of heterozygosity the gametes are also genetically different.

In prophase 1 and metaphase 1 of meiosis individual chromatids pair with the other homologous chromosome with the help of a large protein machinery called the *recombination complex* (see [5, chapter 21]). Tight packing of the four chromatids of the duplicated homologous chromosomes together with the protein machinery and produces a very large *synaptonemal complex*. The processes in the synaptonemal complex may last for days and longer and eventually, prophase 1 ends through disassembling of the synaptonemal complex and initiating metaphase 1. *Chiasmata* being the visual point of crossing-over of chromatid strands that might have occurred already before and during the formation of the complex appear during the phase of disassembly. Resolution of *crossover* in consequence leads to *recombination*. The chromosomes determining the sex of the carrier may behave differently from autosomes. In mammals female sex chromosomes XX behave like autosomes during meiosis. Male sex chromosomes – XY in mammals – however, require special features during meiosis. Although the X and the Y chromosome in a male are not homologous they too must pair and undergo crossover during prophase 1 in order to allow for normal segregation in anaphase 1. Pairing, crossing-over, and segregation are made possible because there are small regions of homology between X and Y at one end or both ends of the chromosomes. The two chromosomes pair and crossover in these regions during prophase 1 and ensure thereby that each sperm cell receives either one X or one Y chromosome – and neither both nor none – and the sperm cells determine whether the zygote develops into a female or male embryo, respectively. Meiosis is regulated differently in female and male mammals. In males meiosis begins begins in sperm precursor cells called *spermatocytes* in the testes at puberty and then goes on continuously. It takes about 24 days for a human spermatocyte to complete meiosis. In females the egg precursor cells, the *oocytes* begin meiosis in the fetal ovary but arrest after the synaptonemal complex has disassembled in metaphase 1. Oocytes complete meiosis only after the female has become sexually mature and the oocyte is released from the ovary during ovulation and the released oocyte completes meiosis only if it is fertilized. In humans some oocytes my be arrested in netaphase 1 for 40 years or more. There are specific stop and start mechanisms in female

meiosis that are lacking in the male. Finally, we remark that according to the current state of knowledge meiosis goes wrong frequently and this leads to early abortion or serious damage of the embryo.

Crossover occurs before the first cell division in meiosis and can be seen in the microscope in the form of a chiasma where a chromatid strands switches from one chromosome to the other. Chiasmata or crossovers are resolved by splitting and ligating DNA and eventually genes from homologous but different chromatids find itself recombined on the same chromosome. As shown in Fig. 2.7 a single crossover is sufficient to produce four different chromosomes. Double and multiple crossovers may occur as well and they give rise to a great variety of gene patterns. How are crossover and recombination related to Gregor Mendel's laws of inheritance? *Linkage equilibrium* in population genetics is achieved when the association of alleles at two or more loci is random. In other words, Mendel made the assumption of random assortment of alleles, which is at the same time the basis for linkage equilibrium and accordingly, every deviation from it is called *linkage disequilibrium*. This deviation can be cast into a quantitative relation. For the sake of simplicity we consider a haplotype[3] for two loci A and B with two alleles each, and the following frequencies for all possible combinations:

$$[\mathsf{A}_1\mathsf{B}_1] = x_{11},\ [\mathsf{A}_1\mathsf{B}_2] = x_{12},\ [\mathsf{A}_2\mathsf{B}_1] = x_{21},\ \text{and}\ [\mathsf{A}_2\mathsf{B}_2] = x_{22}.$$

These frequencies are assumed to be normalized, $\sum_{i,j} x_{ij} = 1$, and this leads to the following frequencies of the alleles:

$$[\mathsf{A}_1] = p_1 = x_{11} + x_{12}\,,\ [\mathsf{A}_2] = p_2 = x_{21} + x_{22}\,,\ \text{and}\ p_1 + p_2 = 1\ ,$$

$$[\mathsf{B}_1] = q_1 = x_{11} + x_{21}\,,\ [\mathsf{B}_2] = q_2 = x_{12} + x_{22}\,,\ \text{and}\ q_1 + q_2 = 1\ .$$

At linkage equilibrium we obtain from trivial statistics $\bar{x}_{ij} = p_i q_j$ and define the linkage disequilibrium by the deviation of the real value from the equilibrium value:

$$D = x_{11} - p_1 q_1 = x_{11}x_{22} - x_{12}x_{21}\ . \tag{2.1}$$

[3]A haplotype in genetics is a combination of alleles at adjacent locations on the chromosome that are transmitted together. A haplotype may be one locus, several loci, or an entire chromosome depending on the number of recombination events that have occurred.

Expressing the state of a population with respect to haplotypes on finds the phrases *"two alleles are in linkage disequilibrium"* for $D \neq 0$, and alternatively *linkage equilibrium* stands for $D = 0$. Because of the various conservation relations linkage disequilibrium is a one-parameter manifold as follows from the table relating haplotype and allele frequencies:

	A_1	A_2	total
B_1	$x_{11} = p_1q_1 + D$	$x_{21} = p_2q_1 - D$	q_1
B_2	$x_{12} = p_1q_2 - D$	$x_{22} = p_2q_2 + D$	q_2
total	p_1	p_2	1

Sometimes the parameter D is normalized

$$\vartheta \;=\; \frac{D}{D_{\max}} \;\text{ with }\; D_{\max} \;=\; \begin{cases} \min\{p_1q_1, p_2q_2\} & \text{if } \; D < 0 \\ \min\{p_1q_2, p_2q_1\} & \text{if } \; D > 0 \end{cases} .$$

As an alternative to ϑ the correlation coefficient between pairs of loci is used (For a comparison of various linkage disequilibrium measures see, e.g., [50]):

$$r \;=\; \frac{D}{\sqrt{p_1p_2q_1q_2}} \; . \tag{2.2}$$

The frequency of recombination between two loci c can be used to demonstrate that linkage disequilibrium converges to zero in absence of other evolutionary factors than Mendelian segregation and random mating. The frequency of the haplotype $\mathsf{A}_1\mathsf{B}_1$ is given by the difference equation

$$x_{11}^{(n+1)} \;=\; (1-c)\, x_{11}^{(n)} \;+\; c\, p_1q_1 \; . \tag{2.3}$$

This equation is readily interpreted: A fraction $(1-c)$ of the haplotypes have not recombined and hence are present in the next generation, multiplication by $x_{11}^{(n)}$ yields the fraction of the not-recombined haplotypes that are $\mathsf{A}_1\mathsf{B}_1$, and a fraction c did recombine the two loci. Random mating presupposed we compute the fraction of the haplotype under consideration: The probability that A_1 is at locus A is p_1 and the probability that the copy has B_1 at locus B. Since the alleles are initially on different loci the events are independent and the probabilities can be simply multiplied. Rewriting of Equ. (2.3) yields

$$x_{11}^{(n+1)} \;-\; p_1q_1 \;=\; (1-c)\,\left(x_{11}^{(n)} \;-\; p_1q_1\right) \;\text{ or }\; D_{n+1} \;=\; (1-c)\, D_n \; ,$$

which for an initial linkage disequilibrium of D_0 takes on the form

$$D_n \;=\; (1-c)^n\, D_0 \;. \tag{2.4}$$

As time progresses and the number of generations approaches infinity we find

$$\lim_{n\to\infty} D_n \;=\; 0 \;,$$

since $\lim_{n\to\infty}(1-c)^n = 0$ because of $0 < 1-c < 1$. After a sufficiently number of generations linkage disequilibrium will disappear due to recombination. The smaller the distance between the two loci, however, the smaller will be the frequency of recombination c, and the slower will be the rate of convergence of D towards zero.

2.2.4 Molecular mechanisms of recombination

The first molecular mechanism of crossover and recombination has been proposed by Robin Holliday in 1964 [150] (For a more recent account see [273]). It centers around the notion of a Holliday junction (Fig. 2.8), which is a covalently linked crossing of two double-helical DNA strands.[4] Holliday junctions combining largely homologous stretches of DNA – as, for example, in case of paired chromatids – can migrate by means of a simple base pair opening and base pair closing mechanism. As shown in the lower part of Fig. 2.8 two DNA double helices become longer and the other two are shortened during migration. According to current knowledge, which is far from satisfactory understanding, homologous crossing-over of chromatid strands is highly regulated with respect to (i) number and (ii) location. There is at least one crossover event between the members of each homolog pairs because this is necessary for normal segregation in metaphase 1, and there is *crossover interference* inhibiting crossover points to be closer than some critical distance. Although the required two strand breaks occurring during meiosis can be located almost everywhere on the chromosome, they are not distributed uniformly: They cluster at *hot spots* where the chromatin is accessible, and

[4] In order to distinguish DNA single strands and double strands as used, for example, in Fig. 2.7 we indicate 5'- and 3'ends of the single strands.

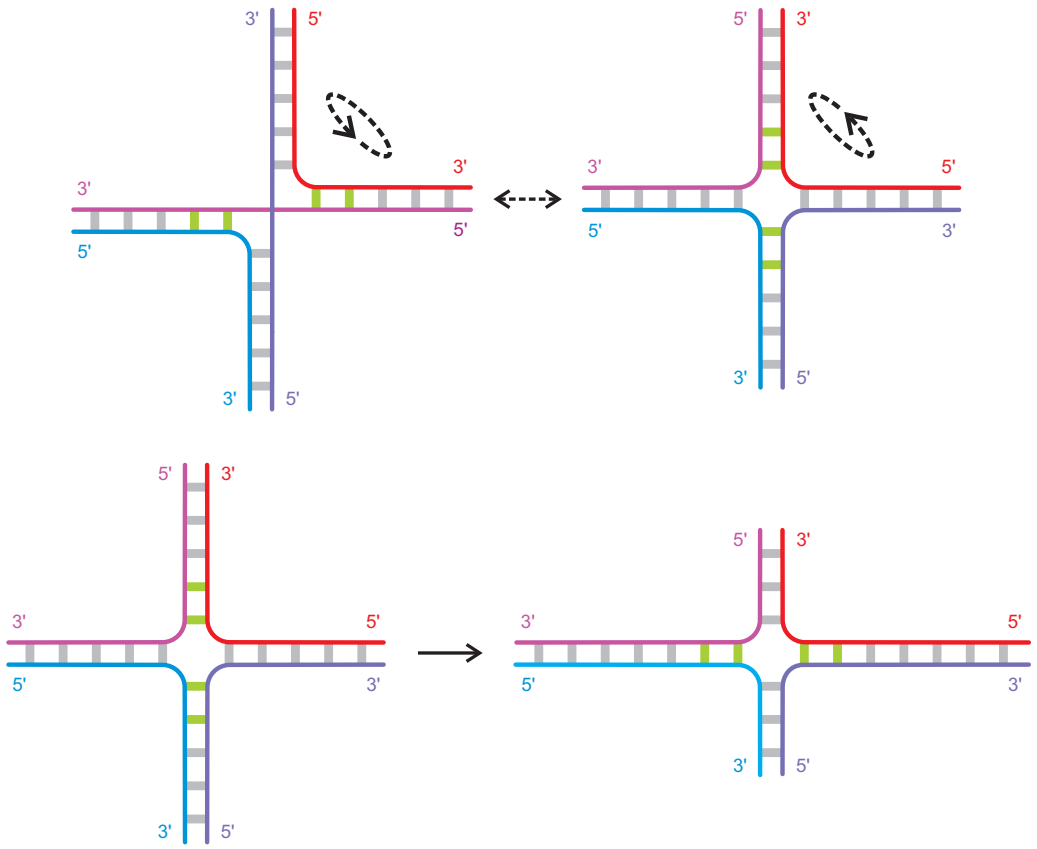


Figure 2.8: **The Holliday junction of two DNA double helices**. At the Holliday junction two stretches of double-helical DNA exchange strands. The upper part of the figure present to schematic views of a Holliday junction, which are interrelated by rotating the upper right part of the sketch by 180° around the diagonal. The lower part shows the base-pair opening and closure mechanism by which Holliday junctions migrate.

they occur only rarely in *cold spots* such as heterochromatin regions around centromeres and telomeres.

The so-called Holliday model for DNA crossover is sketched in Fig. 2.9 [150]. It is initiated by breaks in two DNA single strands called *nicks*, each one situated at one of two aligned DNA molecules. We consider three loci, A, B, and C, with two alleles each, (Aa),(B,b), and (C,c), and the break occurs somewhere in the region between locus A and locus B. In the next

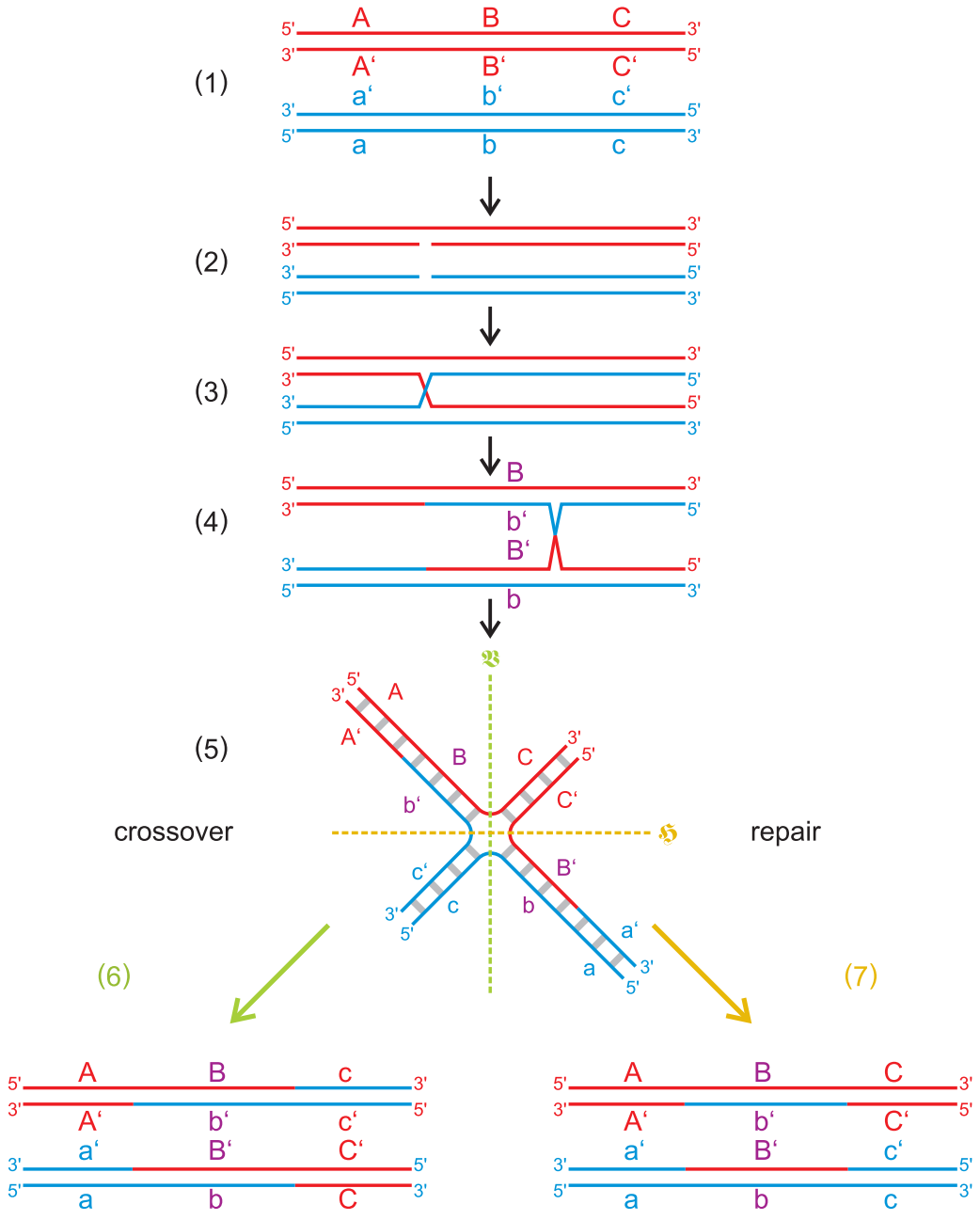


Figure 2.9: **The Holliday model for DNA crossover**. The model shows the formation of a Holliday junction through the repair of two DNA single strand cuts (nicks) and the resolution of the junction resulting in recombination (6; cut along axis $\mathfrak{V}$) or repair (7; cut along $\mathfrak{H}$). Primed letters indicate opposite polarity of the DNA-strand (3'→5'). For details see text.

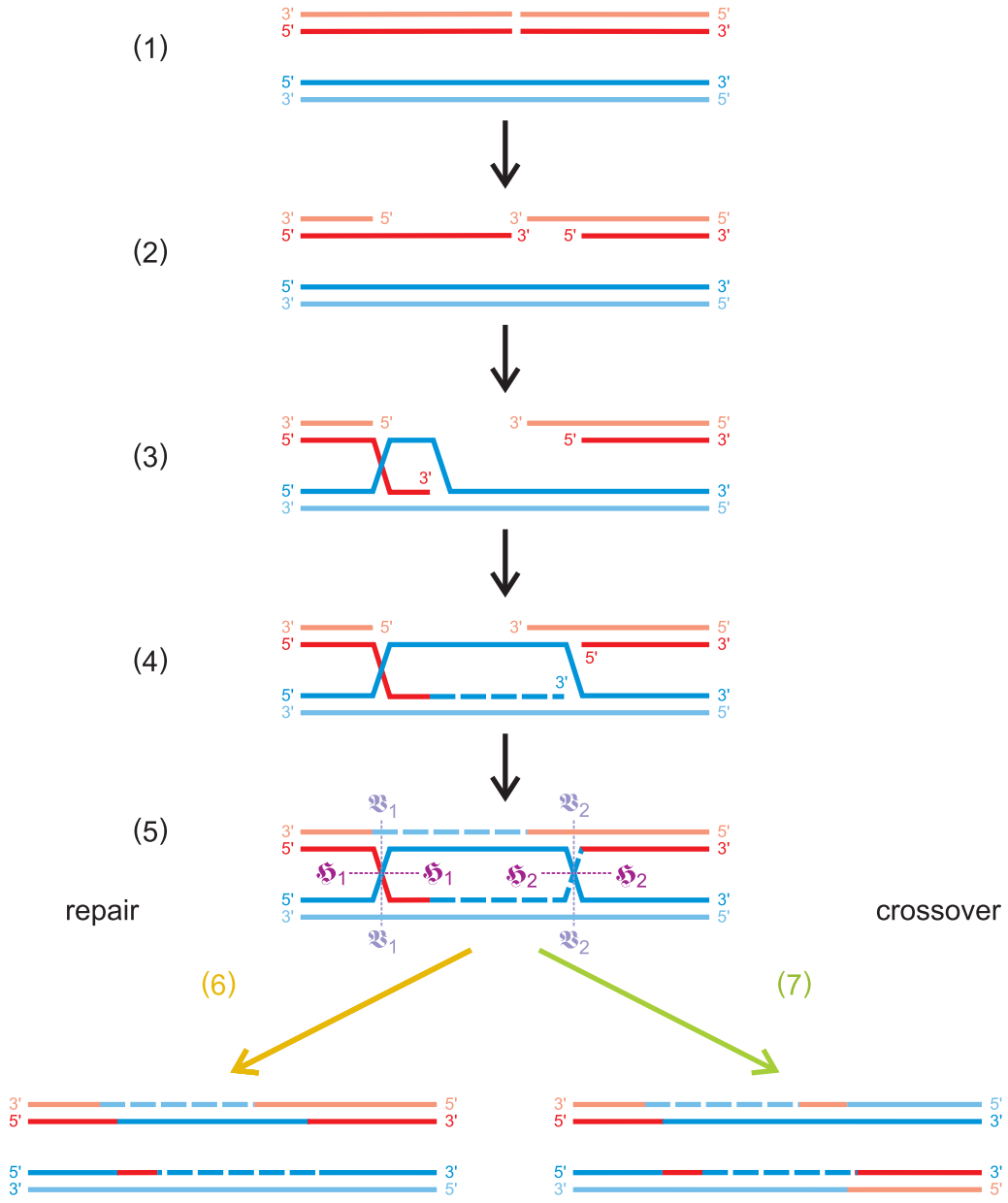


Figure 2.10: **The double strand break repair (DSBR) model for DNA crossover**. The double strand repair model starts from a double strand break and a 5'-exonuclease produces sequence gaps. Strand invasion, DNA-synthesis and second end capture leads to a structure with two Holliday junctions that can be resolved to yield either double strand break repair (resolution at a & b or α & β) or crossover and recombination (resolution at a & β or α & b). Newly synthesized DNA stretches are shown as dashed lines. For details see text.

step the open ends are linked to the other DNA molecules resulting in a Holliday junction. The junction migrates and finds an appropriate site for resolving the crossover somewhere between locus B and locus C. The resulting Holliday junction can be cleaved in two ways, one yielding a recombined DNA with a heteroduplex at locus B and one repaired molecules without recombination but still showing the heteroduplex at B. One of the most important results of the Holiday model was to demonstrate the relation between crossover and DNA damage repair. As outlined in Fig.2.9 (2) the process is initiated by two single strand breaks (nicks), which are closed through the formation of a strand switch from one double helix to the other forming thereby a Holliday junction (3). The Holliday junction migrates along locus B until it reached a point appropriate for resolution (4) and then an enzyme called Holliday junction *resolvase* cuts and ligates the DNA strands eliminating the entanglement of the two DNA double helices (5). The resolution is possible in two different ways: (i) a vertical cut ($\mathfrak{V}$) and (ii) a horizontal cut ($\mathfrak{H}$). The vertical cut resolves the Holliday junction into a structure in which the two chromatid strands show crossover leading to recombination (6) whereas the horizontal cleavage is leading to a structure in which the two nicks have been repaired (7). In both cases the structures differ from the original double helical strands at locus B where they have now heteroduplex pairings Bb. The Holliday-model was complemented ten years after its invention by the more general Meselson-Radding model [208], which starts from a single-strand break in one DNA molecule that becomes the site of strand displacement by a DNA-polymerase. The displaced single-strand pairs with the complementary sequence of a second homologous DNA-molecule and it induces thereby a single strand break in the latter. Migration of the Holliday junction and its resolution is similar to the Holliday-model. One difference between the two models is that the heteroduplex region is confined to one DNA molecule at the beginning in the Meselson-Radding model but is always found in both DNA molecules in the Holliday model.

Work on plasmids in yeast [231] has shown that double strand gap repair can lead to crossing-over but does not always do so. The corresponding *double-strand-break repair* (DSBR) model is sketched in Fig. 2.10. It starts

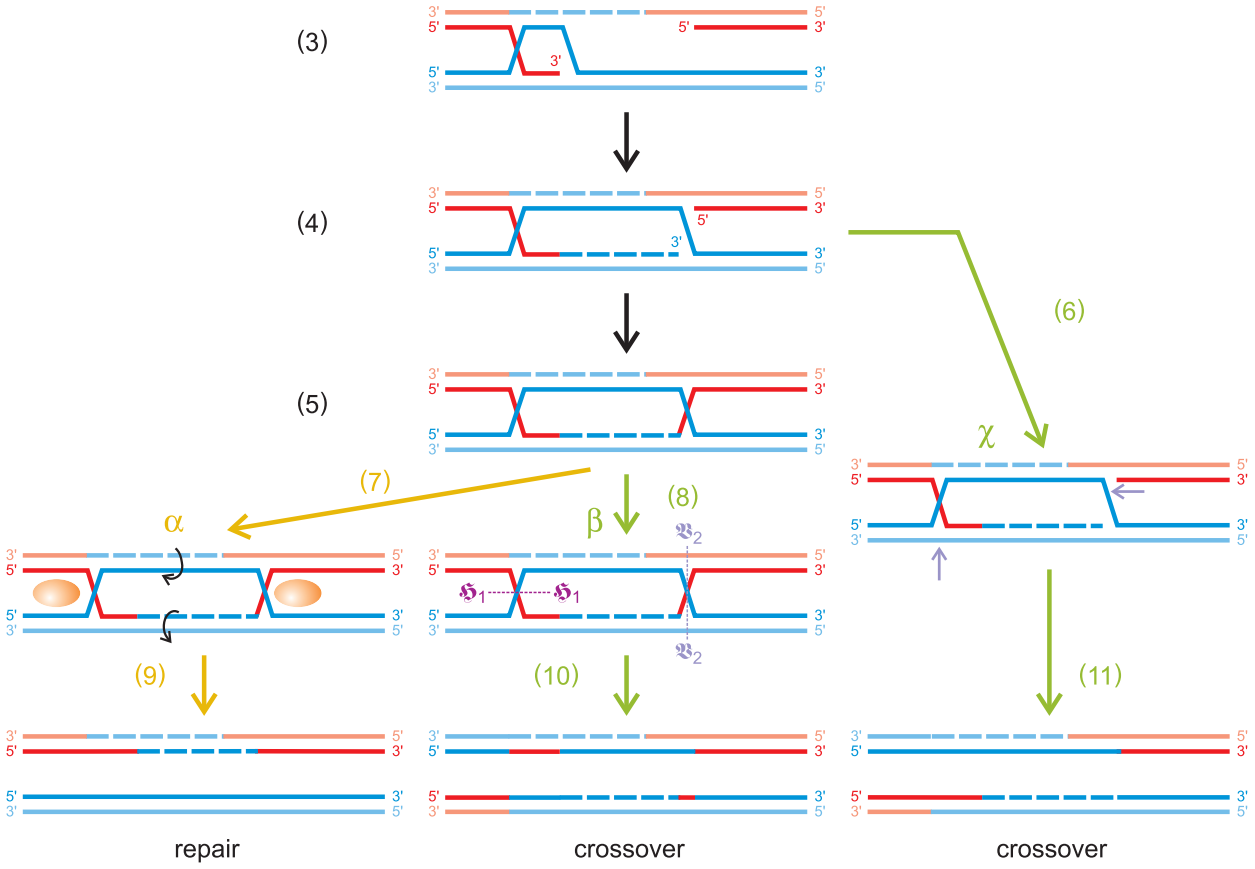


Figure 2.11: **Enzymatic resolution of chiasmata**. Three enzymatic resolution pathways of a double strand break (For stage (1) and (2) see Fig. 2.10). The enzyme complexes are denoted by α, β, and γ, respectively. Newly synthesized DNA stretches are shown as dashed lines. For details see text.

from a double-strand-break (1) and action of 5'-exonucleases as well as some (or no) double strand degradation lead to a structure with free 3'-ends (2). These ends are the active agents in the forthcoming steps: One strand with an active 3'-end invades the double helix of the homologous chromatid and initiates DNA double strand synthesis (3) until the free 3'-end is captured (4). Completion and gap closure on both double helices leads to a structure with two Holliday junctions (5). The resolution of the two junctions leads either to break repair (6) or crossover and recombination (7) depending on the direction of the cuts in the Holliday junctions: Both cuts horizontal or both cuts vertical yields break repair whereas one cut horizontal and one cut vertical produces crossover. Four years after the proposal of the DSBR-model it was tested for recombination of phage λ and plasmid λdv [274, 284]. The most relevant new feature of this paper for the DSBR-model was that a topoisomerase has been suggested to resolve the Holliday junctions. Later topoisomerases were indeed identified that did precisely this job [257, 258].

Research on Holliday junctions and their resolution led to a rather confusing multiplicity of possible pathways and a variety of endonuclease enzymes called *resolvases* were identified in prokaryotes [60], in yeast [23, 230] and eventually also in human cells [37]. In essence, three different pathways for DSBR where found to be most prominent [307]. New light has been shed on the problem when three enzymes from different organisms where shown to promote Holliday junction resolution in analogous manner: the resolvases RuvC from *Escherichia coli*, Yen1 from yeast and GEN1 from human cells [157]. In Fig. 2.11 we show a sketch of the three pathways taken from [281]. The first two steps are identical with the corresponding steps in the DSBR model in Fig. 2.10, one Holliday junction results from DNA strand invasion (3), and then both single strand are completed to full double helices (4). Here we have the first branching point: Either the two open ends are ligated and the second Holliday junction is formed (5) (as in the simple DSBR model) or structure (4) is directly resolved by the protein complex $\chi =$ Mus81-Eme1, which cleaves the asymmetrically [23, 151] and produces a crossover product (11). Two pathways branch out from the double Holliday junction structure (5): One pathway makes use of the protein complex $\alpha =$ BLM-TopoIIIα-

RMI1 and disentangles the structure by two concerted topoisomerase double strand openings and closures (7) [32, 320]. The resulting final product is a double strand repair structure with a heteroduplex region in one double strand (9). The third pathway engages the above mentioned resolvase, $\beta = \text{GEN1}$ or $\beta = \text{Yen1}$, respectively, and resolves the double Holliday junction structure symmetrically by one vertical and one horizontal cut (8) as also shown in Fig. 2.10) leading thereby to crossover (10). Although the 2008 paper [281] had the promising title *Resolving Resolvases: The Final Act?*, research on Holliday junction resolution remained an exciting story until now [217, 279].

After having had a glance on the enormously complicated processes of meiosis and the state of the art in understanding its molecular mechanisms we shall now return to the formal aspects and repeat the basic facts. recombination pure. Homologous chromatid strands do not pair in full length during meiotic prophase 1 but show deviations in the sense that different stretches are aligned to different chromatids, and this leads to chiasmata, crossover and recombination. At least one crossover and recombination event per chromosome is required for successful meiosis, since cells without chiasmata get arrested in metaphase 1 and are eliminated through apoptosis. In essence recombination serves three purposes: (i) repair of double strand breaks that have occurred during replication or pairing, (ii) enabling segregation in metaphase 1 of meiosis, and (iii) creating genetic diversity through recombination.

2.3 Recombination and population genetics

The basic assumption in Mendelian genetics that the genetic information of the parents is split into pieces and recombined in the offspring is introduced by means of a simple relation governing the evolution of genotype distributions for two alleles at a single locus in discrete manner. Fisher's selection equation is an ODE handling an arbitrary number of alleles again at a single locus.

2.4 Hardy-Weinberg equilibrium

The dynamics of recombination is illustrated easily by means of the so called Hardy-Weinberg equilibrium that has been derived independently by Godfrey Harold Hardy [134] and Wilhelm Weinberg [303]. The content of the Hardy-Weinberg equilibrium is the relation between allele frequencies and genotype frequencies in a single locus model. Implicit in the validity of the Hardy-Weinberg equilibrium are ten assumptions, which are often made in population genetics and which we summarize here for clarity [136, p.74]:

(i) organisms are diploid,

(ii) reproduction is sexual,

(iii) generations are discrete and nonoverlapping,

(iv) genes under consideration have two alleles,

(v) allele frequencies are identical in males and females,

(vi) mating partners are chosen at random,

(vii) populations sizes are infinite meaning very large in practice,

(viii) migration is negligible,

(ix) mutation can be ignored, and

(x) natural selection does not affect the alleles under consideration.

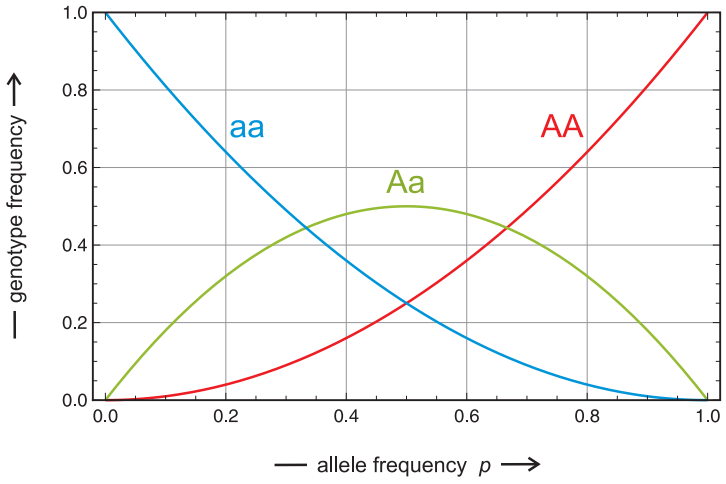


Figure 2.12: **The Hardy-Weinberg equilibrium**. The equilibrium frequencies of the three genotypes, $x = [\mathsf{AA}]$ (red), $y = [\mathsf{Aa}]$ (green), and $z = [\mathsf{aa}]$ (blue), are plotted as a function of the frequency of the dominant allele A, $p = [\mathsf{A}]/([\mathsf{A}] + [\mathsf{a}])$. The frequency of the recessive allele is $q = 1 - p = [\mathsf{a}]/([\mathsf{A}] + [\mathsf{a}])$.

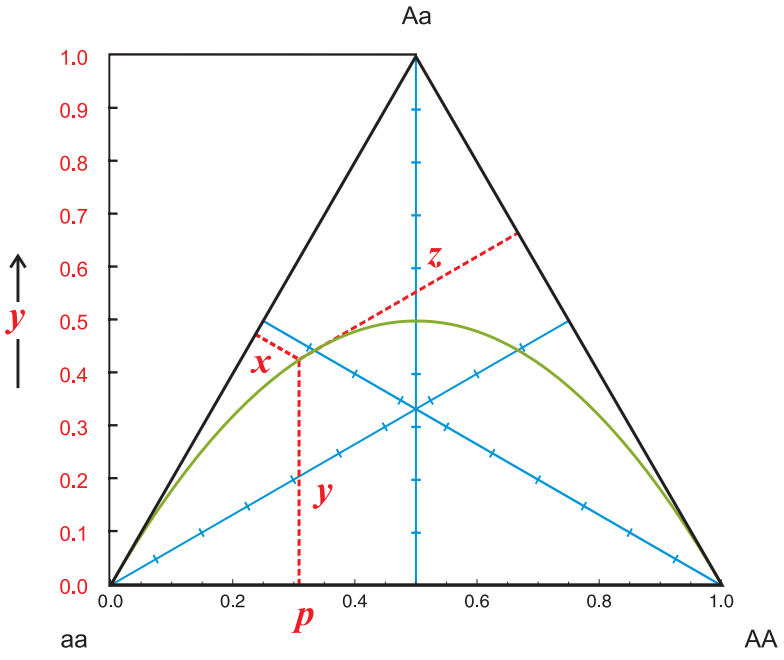


Figure 2.13: **De Finetti illustration of the Hardy-Weinberg equilibrium**. The three genotype frequencies are plotted on a unit simplex $\mathbb{S}_3$: $x + y + z = 1$.

These ten assumptions are often addressed as the Hardy-Weinberg model. In order to derive the relations we assume a diploid population with two alleles A and a where A is dominant and with p and q, $p+q=1$, being the allele frequencies of A and a in the population. At Hardy-Weinberg equilibrium we obtain the three genotype frequencies

$$x \,=\, [\mathsf{AA}] \,=\, p^2\,,\; y \,=\, [\mathsf{Aa}] \,=\, 2\,pq\,,\; z \,=\, [\mathsf{aa}] \,=\, q^2\;. \tag{2.5}$$

In order to show the one-step convergence towards the equilibrium relations we start from an initial population Υ_0 with a distribution of genotypes (AA),(Aa), and (aa) being $x:y:z=p_0:2q_0:r_0$, respectively, fulfilling the condition $p_0+2q_0+r_0=1$. The sum is now written as $(p_0+q_0)+(q_0+r_0)$, we build the square of both sides and find:

$$p_1 \,+\, 2q_1 \,+\, r_1 \;=\; (p_0+q_0)^2 \,+\, 2\,(p_0+q_0)(q_0+r_0) \,+\, (q_0+r_0)^2 \;=\; 1\;.$$

The individual frequencies are $p_1=(p_0+q_0)^2$, $q_1=(p_0+q_0)(q_0+r_0)$ and $r_1=(q_0+r_0)^2$, which is already Hardy's result for random mating (2.5). The *equivalence condition* is readily verified:

$$E_1 \;=\; q_1^2 - p_1\,r_1 \;=\Big((p_0+q_0)(q_0+r_0)\Big)^2 - (p_0+q_0)^2(q_0+r_0)^2 \;=\; 0\;. \tag{2.6}$$

It is straightforward to show now that for all generations after the first one the Hardy-Weinberg equilibrium is fulfilled:

$$\begin{aligned} p_2 \;&=\; (p_1+q_1)^2 \;=\; p_1^2+2p_1q_1+q_1^2 \;= \\ &= (p_0+q_0)^2\left((p_0+q_0)^2+2(p_0+q_0)(q_0+r_0)+(q_0+r_0)^2\right) \;=\; (p_0+q_0)^2\;. \end{aligned}$$

Accordingly $p_2=p_1$ and this remains so in all succeeding generations. The same holds for the other two genotypes.

The generalization of the Hardy-Weinberg equilibrium to n alleles is straightforward: Assume a distribution $(p_1,p_2,\ldots,p_n)$ for n alleles $\mathsf{A}_1,\mathsf{A}_2\ldots\mathsf{A}_n$ with $\sum_{i=1}^n p_i=1$ and then the Hardy-Weinberg equilibrium is achieved when the genotype frequencies fulfil

$$x_i \;=\; [\mathsf{A}_i\mathsf{A}_i] \;=\; p_i^2 \;\text{ and }\; y_{ij} \;=\; [\mathsf{A}_i\mathsf{A}_j] \;=\; 2\,p_ip_j\,;\; i,j=1,2,\ldots,n\;. \tag{2.7}$$

This equation will be the basis for Fisher's selection equation, which will be discussed in the next section 2.5. Another straightforward generalization the case of polyploidy that leads to a genotype distribution according to the binomial distribution for tetraploidy and two alleles, A and a, with the frequencies (p, q) we find

$$[\mathsf{AAAA}] = p^4\,,\ [\mathsf{AAAa}] = 4p^3q\,,\ [\mathsf{AAaa}] = 6p^2q^2\,,\ [\mathsf{Aaaa}] = 4pq^3\,,\ [\mathsf{aaaa}] = q^4\,.$$

The derivation of the completely general form, n alleles and m-ploidy, is left to the reader as an exercise.

As an example of Hardy Weinberg equilibrium in case of dominance we consider the human Rh blood groups (see section 2.1). The dominant allele D codes for the Rhesus antigen, which is presented on the surface of red blood cells, whereas the d allele fails to code for the antigen. Accordingly, the two genotypes DD and Dd unfold the Rhesus positive phenotype (Rh^+), whereas dd has the Rhesus negative phenotype (Rh^-). The frequency of Rh^+ phenotypes among American Caucasians is about 0.858, leaving 14.2 % for Rh^- people [216]. Left with this information only the data are insufficient to calculate the genotype frequencies, because there is no way to distinguish between DD and Dd since both give rise to the Rh^+ phenotype. Under the assumption of random mating, however, the relative proportions of DD and Dd genotypes are given by the Hardy-Weinberg principle: The genotype frequencies at equilibrium are given by p^2, $2pq$, and q^2, respectively. An estimate of q from the known frequency of the homozygote $[\mathsf{dd}] = q^2$ is straightforward:[5] From $\hat{q}^2 = 0.142$ we obtain $\hat{q} \approx \sqrt{0.142} = 0.3768$. The result is easily generalized: If R is the frequency of homozygous recessive genotypes in a population of N individuals, then $\hat{q}$ and its standard deviation σ[6] are obtained from

$$\hat{q} \;=\; \sqrt{R} \;\text{ and }\; \sigma(\hat{q}) \;=\; \sqrt{\frac{1-R}{4N}}\,. \tag{2.8}$$

[5]The remark *estimate* refers to the uncertainty, how well the assumption of random mating is fulfilled. Accordingly, we denote the estimated values for the allele frequencies by $\hat{p}$ and $\hat{q}$ in order to distinguish them from the exact values p and q, respectively.

[6]The standard deviation is calculated under the assumption of a binomial distribution in the limit of the normal distribution: $\sigma = \sqrt{\hat{p}\,\hat{q}/N} = \sqrt{(1-\hat{q})\,\hat{q}/n}$.

From $\hat{q}$ we obtain $\hat{p} = 1 - \hat{q} = 0.6232$ and we calculate for the genotype frequencies: $[\mathsf{DD}] = \hat{p}^2 = 0.3884$, $[\mathsf{Dd}] = 2\,\hat{p}\hat{q} = 0.4696$, and $[\mathsf{dd}] = \hat{q}^2 = 0.1420$, and 54.7 % of the Rh^+ people are heterozygous. No χ^2-test is possible in this case because there are zero degrees of freedom (subsection 2.6.1).

It is remarkable that the Hardy-Weinberg principle is considered as the basis of many models in population genetics, although on cannot assume that the ten conditions listed above a fulfilled in reality. Many test for random mating (item vi) based on deviations from the Hardy-Weinberg equilibrium have been developed, for example [76, 129], and it was shown that deviations are quite common. Sten Wahlund had shown that a subdivision of a population in subpopulations leads to a reduction of heterozygosity [299]. Even when the subpopulations are in Hardy-Weinberg equilibrium, the total population is not. Another critical point is the absence of effects of natural selection (item x), which is a typical idealization and very hard to check independently of deviations from Hardy-Weinberg equilibrium. Therefore it is advisable to consider the Hardy-Weinberg formula as a reference state and to analyze deviations as consequences of the lack of validity of the basic assumptions. We shall come back to the delicate problem of generality, explanatory adequacy, scope, and applicability in the next chapter 3.

2.5 Fisher's selection equation and the fundamental theorem

Here, we present only the continuous time approach more common stochastic models with discrete generations will be discussed in chapter 8. In order to study the process of selection among n alleles at a single locus under random mating and recombination Ronald Fisher's [92] conceived a selection equation for alleles:

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = \sum_{i=1}^{n} a_{ji} x_j x_i - x_j \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ik} x_i x_k = x_j \left(\sum_{i=1}^{n} a_{ji} x_i - \phi \right) \tag{2.9}$$

$$\text{with} \quad \phi = \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ik} x_i x_k \,. \tag{2.10}$$

The variables x_j are the allele frequencies in the population. The two conditions $a_{ij} > 0 \,\forall\, i, j = 1, 2, \ldots, n$ and $x_i \geq 0 \,\forall\, i = 1, 2, \ldots, n$ will guarantee

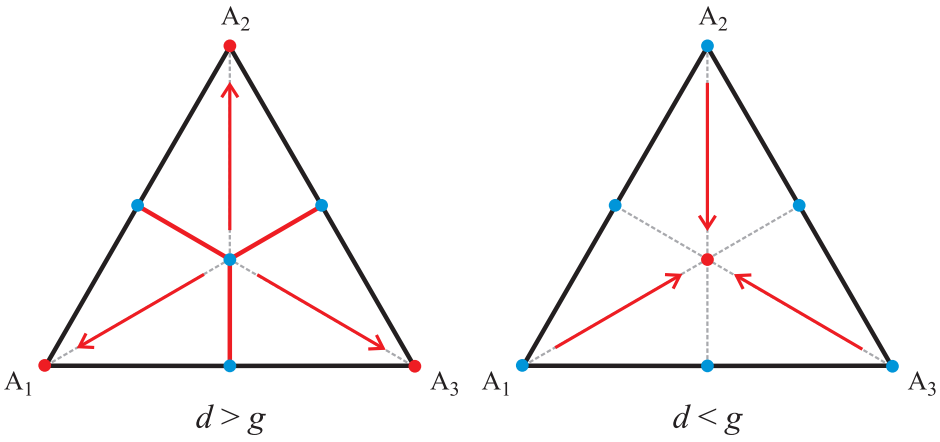


Figure 2.14: **Selection dynamics of three alleles at one locus**. Selection is shown in the space of normalized allele concentrations, $\sum_{i=1}^{3} x_i = 1$, which is the unit simplex $\mathbb{S}_3$. In the general case, the dynamical system has seven stationary points. For the sake of simplicity we consider the symmetric case with equal diagonal and equal off diagonal elements of the parameter matrix A. If the diagonal elements dominate, $d > g$, all three corners represent asymptotically stable states ($x_1 = 1$, $x_2 = 1$, and $x_3 = 1$). For larger off-diagonal elements, $d < g$, the only asymptotically stable state is the center of the simplex, $x_1 = x_2 = x_3 = 1/3$. Color code: asymptotically stable states in red, saddle points and sources in blue.

$\phi(t) \geq 0$. Summation of allele frequencies, $\sum_{i=1}^{n} x_i(t) = c(t)$, yields again an equation for $\mathrm{d}c/\,\mathrm{d}t$ that is identical to (1.10) and hence, the population is confined again to the unit simplex for $\sum_{i=1}^{n} x_i(0) = 1$. The rate parameters a_{ij} form a quadratic matrix

$$\mathrm{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix} .$$

The dynamics of equ. (2.9) may be very complicated for general matrices A and involve oscillations as well as deterministic chaos [243, 253]. In case of Fisher's selection equation, however, we are dealing with a symmetric matrix for biological reasons,[7] and then the differential equation can be subjected to straightforward qualitative analysis.

Qualitative analysis of equ. (2.9) yields $2^n - 1$ stationary points, which depending on the elements of matrix A may lie in the interior, on the boundary or outside the unit simplex $\mathbb{S}_n^{(1)}$. In particular, we find at maximum one equilibrium point on the simplex and one on each subsimplex of the boundary. For example, each corner, represented by the unit vector $\boldsymbol{e}_k = \{\bar{x}_k = 1, x_i = 0 \,\forall\, i \neq k\}$, is a stable or unstable stationary point. In case there is an equilibrium in the interior of $\mathbb{S}_n^{(1)}$ it may be stable or unstable depending on the elements of A. In summary, this leads to a rich collection of different dynamical scenarios, which share the absence of oscillations or chaotic dynamics.

The coordinates of the stationary points are derived through solution of the equations derived from (2.9) by putting $dx_i/dt = 0$ for $i = j, k, l$ with $\bar{x}_j + \bar{x}_k + \bar{x}_l = 1$ and $x_i = 0 \,\forall\, i \notin (j, k, l)$:

$$
\begin{aligned}
&\text{corner} \quad j: && \bar{x}_j = 1 \\
&\text{edge} \quad \overline{jk}: && \bar{x}_j = \frac{a_{kk} - a_{jk}}{a_{jj} - 2a_{jk} + a_{kk}}\,,\ \bar{x}_k = \frac{a_{jj} - a_{jk}}{a_{jj} - 2a_{jk} + a_{kk}} \\
&\text{face} \ \triangle jkl: && \bar{x}_j = \frac{Z_j}{D}\,,\ \bar{x}_k = \frac{Z_k}{D}\,,\ \bar{x}_l = \frac{Z_l}{D} \ \text{ with} \\
& && Z_j = a_{jl}a_{kk} + a_{jk}a_{ll} + a_{kl}^2 - a_{jk}a_{kl} - a_{jl}a_{kl} - a_{kk}a_{ll}\,, \\
& && Z_k = a_{jj}a_{kl} + a_{jl}^2 + a_{kj}a_{ll} - a_{jl}a_{jk} - a_{kl}a_{jl} - a_{jj}a_{ll}\,, \\
& && Z_l = a_{jk}^2 + a_{jl}a_{kk} + a_{jj}a_{kl} - a_{jk}a_{jl} - a_{jk}a_{kl} - a_{jj}a_{kk}\,, \\
& && D = Z_j + Z_k + Z_l\,.
\end{aligned}
$$

[7] Fisher's equation is based on the assumption that phenotypes are insensitive to the origin of the parental alleles on chromosomes. Phenotypes derived from genotype $\mathbf{A}_i\mathbf{A}_j$ are assumed to develop the same properties, no matter which allele, , on the chromosomal locus comes from the mother and which comes from the father. New results on human genetic diseases have shown, however, that this assumption can be questioned.

Local stability analysis on the simplex S_3 through diagonalization of the (2×2) Jacobian yields for the corners

$$\text{corner} \quad j: \qquad \lambda_1 = a_{jk} - a_{jj}\,,\ \lambda_2 = a_{jl} - a_{jj}\ ,$$

and for the edges

$$\text{edge} \quad \overline{jk}: \qquad \lambda_1 = \frac{(a_{jk} - a_{jj})(a_{jk} - a_{kk})}{a_{jj} - 2a_{jk} + a_{kk}}\,,$$

$$\lambda_2 = -\frac{a_{jj}a_{kk} - a_{jj}a_{kl} - a_{jl}a_{kk} - a_{jk}^2 + a_{jk}a_{jl} + a_{jk}a_{kl}}{a_{jj} - 2a_{jk} + a_{kk}}\,.$$

The corner j is asymptotically stable for $(a_{jk}, a_{jl}) < a_{jj}$ or in other words, if the homozygote $\mathbf{A}_j\mathbf{A}_j$ has higher fitness than the two heterozygotes $\mathbf{A}_j\mathbf{A}_k$ and $\mathbf{A}_j\mathbf{A}_l$. The stationary point on the edge $\overline{jk}$ is unstable for $(a_{jj}, a_{kk}) > a_{jk}$ because λ_1 is positive. We dispense here for a more detailed discussion of the eventually quite sophisticated situation and refer to the simplified model discussed below.

The calculation of the eigenvalues of the Jacobian at the stationary point in the interior of the face $\triangle jkl$ is even more involved but, nevertheless, can be computed analytically. We present the results for the stationary point on the face $\triangle jkl$ in order to demonstrate the strength and the limits of machine based symbolic computation. For the two eigenvalues of the Jacobian we find:

$$\begin{aligned}
\lambda_{1,2} &= (Q_1 \pm \sqrt{Q_2})/(2D^2) \ \text{ with} \\
Q_1 &= (a_{kl} - a_{kk})\,P_1 + (a_{jl} - a_{jj})\,P_2 \ \text{ and} \\
Q_2 &= \Big((a_{jl} - a_{jj})(a_{kl} - a_{kk}) - (a_{jk} - a_{jl})(a_{jk} - a_{kl})\Big)P_1P_2\ , \\
P_1 &= a_{jk}a_{jl} - a_{jk}a_{ll} - a_{jj}a_{kl} + a_{jl}a_{kl} + a_{jj}a_{ll} - a_{jl}^2\ , \\
P_2 &= a_{jk}a_{kl} - a_{jk}a_{ll} - a_{jl}a_{kk} + a_{jl}a_{kl} + a_{kk}a_{ll} - a_{kl}^2\ , \\
D &= a_{jk}^2 + a_{jl}^2 + a_{kl}^2 - a_{jj}a_{kk} - a_{jj}a_{ll} - a_{kk}a_{ll} + \\
&\quad + 2(a_{jj}a_{kl} + a_{jl}a_{kk} + a_{jk}a_{ll} - a_{jk}a_{jl} - a_{jk}a_{kl} - a_{jl}a_{kl})
\end{aligned}$$

Q_2 when completely expanded becomes a sixth order polynomial in a_{jk} with more than 120 terms. Although fully analytical the expressions are pro-

hibitive for further calculations and will be studied numerically in practical calculations.

A simple example of a three-allele one-locus model is shown on Fig. 2.14: The matrix of rate parameters is simplified to

$$\mathrm{A} \;=\; \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \;=\; \begin{pmatrix} d & g & g \\ g & d & g \\ g & g & d \end{pmatrix} \;,$$

which has only two parameters the diagonal terms d representing the fitness of the homozygotes, and g the off-diagonal elements for the heterozygotes. From $a_{11} = a_{22} = a_{33} = d$ and $a_{12} = a_{13} = a_{23} = g$ we have $Q_2 = 0$. In this fully symmetric case the coordinates of stationary points and the eigenvalues of the Jacobian fulfil very simple expressions:

$$\begin{aligned} &\text{corner} \quad (1,0,0): && \lambda_1 = \lambda_2 = d - f \\ &\text{edge} \quad (\frac{1}{2},\frac{1}{2},0): && \lambda_1 = -\lambda_2 = -\frac{1}{2}(g-d) \\ &\text{face} \quad (\frac{1}{3},\frac{1}{3},\frac{1}{3}): && \lambda_1 = \lambda_2 = -\frac{1}{3}(g-d)\,. \end{aligned}$$

The critical quantity here is the difference between the off-diagonal and the diagonal term of matrix A, $g-d$. As long as $d > g$ is fulfilled – corresponding to higher fitness of homozygotes, $\mathbf{A}_j\mathbf{A}_j$ and $\mathbf{A}_i\mathbf{A}_i$, than heterozygotes, $\mathbf{A}_j\mathbf{A}_i$ and $\mathbf{A}_i\mathbf{A}_j$ – the corners are stable stationary points and depending on initial conditions one allele, $\mathbf{A}_j$ or $\mathbf{A}_i$, is selected. For *overdominance* of heterozygotes, $g > d$, the stable points are on the edges or in the interior of the face (Fig. 2.14). Multiple stationary states do occur and more than one may be stable and the outcome of population dynamics need not be uniquely defined. Instead depending on initial conditions the distribution of alleles may approach one of the local optima [4, 86, 147, 245].

In order to analyze the behavior of the mean fitness $\phi(t)$ we introduce mean rate parameters $\overline{a}_i = \sum_{j=1}^{n} a_{ij}x_j$, which facilitate the forthcoming anal-

ysis. The time dependence of ϕ is now given by

$$\begin{aligned}
\frac{d\phi}{dt} &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\left(\frac{dx_i}{dt}\cdot x_j + x_i\cdot\frac{dx_j}{dt}\right) = 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ji}\cdot x_i\cdot\frac{dx_j}{dt} = \\
&= 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ji}\cdot x_i\left(\sum_{k=1}^{n} a_{jk}x_jx_k - x_j\sum_{k=1}^{n}\sum_{\ell=1}^{n} a_{k\ell}x_kx_\ell\right) = \\
&= 2\sum_{j=1}^{n} x_j\sum_{i=1}^{n} a_{ji}x_i\sum_{k=1}^{n} a_{jk}x_k - 2\sum_{j=1}^{n} x_j\sum_{i=1}^{n} a_{ji}x_i\sum_{k=1}^{n} x_k\sum_{\ell=1}^{n} a_{k\ell}x_\ell = \\
&= 2\sum_{j=1}^{n} x_j\,\overline{a}_j^2 - 2\sum_{j=1}^{n} x_j\,\overline{a}_j\sum_{k=1}^{n} x_k\,\overline{a}_k = \\
&= 2\left(<\overline{a}^2> - <\overline{a}>^2\right) = 2\,\mathrm{var}\{\overline{a}\} \geq 0\,. \qquad (2.11)
\end{aligned}$$

Again we see that the flux $\phi(t)$ is a non-decreasing function of time, and it approaches an optimal value on the simplex. This result is often called Fisher's fundamental theorem of evolution (see, e.g., [86]). As said above, multiple stationary states do occur and more than one may be stable. This implies that the optimum, $\phi(t)$ is approaching, need not be uniquely defined. Instead $\phi(t)$ may approach one of the local optima and then the outcome of the selection process will depend on initial conditions [4, 86, 147, 245].

Three final remarks are important for the naïve interpretation of Fisher's fundamental theorem: (i) Selection in the one-locus system when it follows Equ. (2.9) optimizes mean fitness of the population, (ii) the outcome of the process need not be unique since the mean fitness ϕ may have several local optima on the unit simplex, and (iii) optimization behavior that is susceptible to rigorous proof is restricted to the one locus model since systems with two or more gene loci may show different behavior of $\phi(t)$. In particular, epistasis, linkage disequilibrium, and frequency dependent fitness may lead to situations in which the mean fitness is decreasing [214]. The conventional opinion considered the fundamental theorem as wrong and a mistake of Ronald Fisher who – by the way – made so many other important contributions to mathematics, statistics, and population genetics.

Recalling Fisher's own verbal formulation of the fundamental theorem: *The rate of increase of the mean fitness of any species at any time is equal*

to its genetic variation at that time, allow for many different interpretations. A more recent reinterpretation the theorem [103, 228] tries to rehabilitate Fisher's claim for universality. These interpretations are based on partitioning genetic variation in additive and nonadditive, in other words epistatic, effects and intrinsic and environment caused contributions. The misunderstanding of Fisher's original intensions is mainly attributed to the thirty years controversy about the meaning of natural selection between Fisher and Wright. We shall come back to these differences in view when we discuss Fisher's theorem and Wright's concept of adaptive landscapes in the light of a general theory of evolution (chapter 3).

2.6 Evaluation of data and the Fisher-Mendel controversy

Derivation and experimental verification of statistical laws requires a firm mathematical theory of testing whether or not the observed regularity follows significantly from the harvested data. Ronald Fisher criticized Mendel's work [93] and brought up the argument that Mendel's data are *too good* and, presumably, were slightly manipulated by the author. Fisher's paper was the beginning of a seventy years long debate and eventually led to a monograph stating that it is high-time to end the controversy [104]. It is fair to say that Mendel work in essence has been rehabilitated but Fisher's statistical perfection in data analysis created a new standard in th e validation of statistical laws. We dispense from here from all historical details but use the Mendel-Fisher discussion to digress on statistical methods that allow for an evaluation of the statistical significance of harvested data.

2.6.1 The χ^2-distribution

The conventional statistical test for data from random sampling is called Pearson's χ^2-test, because it has been introduced by the statistician Karl Pearson [233] and assesses two kinds of comparisons: (i) a quality test that establishes whether or not an observed value distribution differs from a theoretical distribution, and (ii) a test of independence that assesses whether paired observations on two variables are independent of each other.

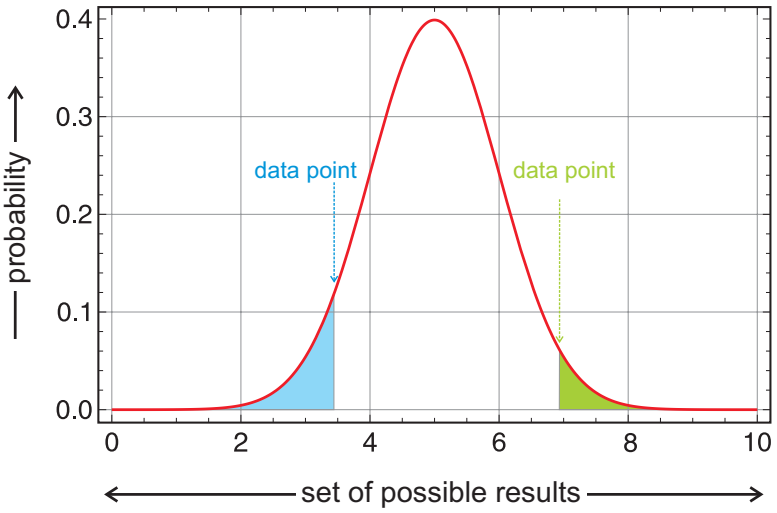


Figure 2.15: **The p-value in significance test of null hypothesis**. The figure shows the definition of the p-value. The bell-shaped curve (red) is the probability density function (PDF) of possible results. Two specific data points are shown one at values above the most frequent outcome at $x = 5$ near $x = 7$ (green) and the other one at $x \approx 3.5$ (blue). The p-value – not to be mistaken for a score – is the cumulative probability of more extreme cases, i.e., results that are further away of the most frequent outcome than the data point and obtained as the integral under the PDF. Depending of the position of the observed result this integral has to be taken to higher (green) or lower (blue) values of x, respectively.

The test is based on the χ^2-distribution with the following properties:[8] $\chi^2_k(x)$ is a one-parameter probability distribution, which is defined on the positive real axis (support: $x \in [0, +\infty)$), and which is defined as the sum of the squares of k independent random variables $\mathcal{Z}_1, \ldots, \mathcal{Z}_k$ that fulfil the standard normal distribution:

$$\mathcal{Q} \;=\; \sum_{i=1}^{k} \mathcal{Z}_i^2 \;,$$

The parameter k is the number of the degrees of freedom. The distribution

[8] In case it is important to distinguish between the test statistics and the χ^2-distribution the test is named Pearson X^2 test.

of $\mathcal{Q}$ is given by

$$\begin{aligned} \text{probability density function:} \quad f(x;k) &= \frac{1}{2^{\frac{k}{2}}\,\Gamma\left(\frac{k}{2}\right)}\, x^{\frac{k}{2}-1}\, e^{-\frac{x}{2}} \ , \\ \text{cumulative distribution function:} \quad F(x;k) &= \frac{1}{\Gamma\left(\frac{k}{2}\right)}\, \gamma\left(\frac{k}{2}, \frac{x}{2}\right) \ , \end{aligned} \tag{2.12}$$

whereby $\chi^2_k(x) = f(x;k)$. The functions $\Gamma(z)$, $\gamma(s,z)$ and $\Gamma(s,z)$ are the Gamma function and the lower and the upper incomplete Gamma function, respectively:

$$\begin{aligned} \Gamma(z) &= \int_0^\infty e^{-t}\, t^{z-1}\, \mathrm{dt} \quad \text{and} \quad \Gamma(n) = (n-1)!; \ , n \in \mathbb{N}_{\not\vdash} \ , \\ \gamma(s,z) &= \int_0^z e^{-t}\, t^{z-1}\, \mathrm{dt}\,, \ \text{and} \\ \Gamma(s,z) &= \int_z^\infty e^{-t}\, t^{z-1}\, \mathrm{dt} \ . \end{aligned}$$

The χ^2 density and distribution functions are available in tables that were used before the ubiquitous availability of computers, which allow for straightforward calculation of numerical values from the functions available in all statistics packages. Since $\Gamma(1) = 1$ the case $k = 2$ is particularly simple and serves as an example:

$$f(x;2) = \frac{1}{2}\, e^{-\frac{x}{2}} \quad \text{and} \quad F(x;2) = \int_0^x f(x;2)\, \mathrm{d}x = 1 - e^{-\frac{x}{2}} \ .$$

The integration starts here at $x = 0$ rather than at $x = -\infty$ because the support of the distribution is restricted to the positive real axis.

In order to perform a specific test the start is to define the null hypothesis that is the assumption of a theoretical distribution of the measured values commonly in form of an expected partitioning of N observations in n cells. In case of a discrete uniform distribution as null hypothesis – as it is very often the case – the theoretical frequency is given by

$$\varepsilon_i = \frac{N}{n}\,, \ i = 1, 2, \ldots, n \ .$$

Other common null hypotheses are the assumption of a normal, a Poisson or a binomial distribution. Next the test statistic is calculated according to

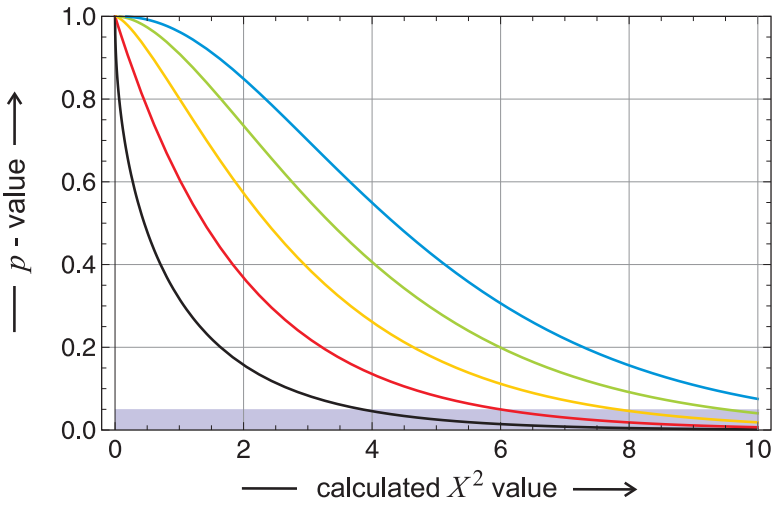


Figure 2.16: **Calculation of the p-value in significance test of null hypothesis**. The figure shows the p-values from Equ. (2.14) as a function of the calculated values of X^2 for the d-values 1 (black), 2 (red), 3 (yellow), 4 (green), and 5 (blue). The highlighted area at the bottom of the figure shows the range where the null hypothesis is rejected.

Pearson's cumulative test statistic

$$X^2 = \sum_{i=1}^{n} \frac{(\nu_i - \varepsilon_i)^2}{\varepsilon_i} , \tag{2.13}$$

where ν_i is the number of observations that were falling into cell $\mathcal{C}_i$. The cumulative test statistic X^2 converges to the χ^2 distribution in the limit $N \to \infty$ – just as a mean value of a stochastic variable, $\overline{\mathcal{Z}} = \sum_{i=1}^{N} z_i$ converges to the expectation value $\lim_{N\to\infty} \overline{\mathcal{Z}} = E\{\mathcal{Z}\}$. This implies that X^2 is never exactly equal to χ^2 and the approximation that will always become better when the sample size is increased. Usually a lower limit is defined for the number of entries in the cells to be considered, values between 5 and 10 are common.

Now 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, n, represents the maximal number of degrees of freedom, which is reduced

by $\pi = s + 1$ where s is the number of parameters used in fitting of the distribution of the distribution of the null hypothesis and accordingly we have $d = n - \pi$. Considering the uniform distribution that is parameter free we find $d = n - 1$ and this is readily interpreted: Because the data fulfil $\sum_{i=1}^{n} \pi_i = N$ only $n - 1$ cells can be filled independently. Eventually, we determine the p-value of the data sample as a measure of statistical significance. Precisely, the p-value is the probability of obtaining a test statistic that is at least as extreme as the actually observed one under the assumption that the null hypothesis is true. We call a value **a** *more extreme* than **b** if **a** is less likely to occur under the null hypothesis as **b**. As shown in Fig. 2.15 this probability is obtained as the integral below the PDF from the calculated X^2-value to $+\infty$. In case of the χ^2 distribution we have

$$p \;=\; \int_{X^2}^{+\infty} \chi_d^2(x)\,\mathrm{d}x \;=\; 1 \;-\; \int_0^{X^2} \chi_d^2(x)\,\mathrm{d}x \;= 1 \;-\; F(X^2; d)\;, \qquad (2.14)$$

which involves the cumulative distribution function $F(x; d)$ defined in Equ. (2.12). Commonly, the null hypothesis is rejected when p is smaller than the significance level: $p < \alpha$ with $0.001 \leq \alpha \leq 0.05$. If the condition $p < \alpha$ is fulfilled one says the null hypothesis is statistically significantly rejected.

A simple example is used for the purpose of illustration: Two random samples of N animals was drawn from a population, ν_1 were males and ν_2 were females with $\nu_1 + \nu_2 = N$. The first sample,

$$N = 322,\, \nu_1 = 170,\, \nu_2 = 152 : X^2 = \frac{(170 - 161)^2 + (152 - 161)^2}{322} = 0.503\,,$$
$$p = 1 - F(0.503; 1) = 0.478\,,$$

clearly supports the null hypothesis that that males and females are equally frequent since $p > \alpha \approx 0.05$. The second sample,

$$N = 467,\, \nu_1 = 207,\, \nu_2 = 260 : X^2 = \frac{(207 - 233.5)^2 + (260 - 233.5)^2}{233.5} = 6.015\,,$$
$$p = 1 - F(6.015; 1) = 0.0142\,,$$

leads to a p-value, which definitely is at the lower limit or below the critical limit and the rejection of the null hypothesis is statistically significant.

The test of independence is relevant for situations when an observation 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 2.6.2). 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

$$\varepsilon_{ij} \;=\; \frac{\sum_{k=1}^{n}\nu_{ik}\,\sum_{k=1}^{m}\nu_{kj}}{N}\;, \tag{2.15}$$

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

$$X^2 \;=\; \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(\nu_{ij}-\varepsilon_{ij})^2}{\varepsilon_{ij}}\;. \tag{2.16}$$

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 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, $p = 1 - F(X^2; d)$, and a value of p less than a predefined critical value, commonly $p < 0.05$, 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.

2.6.2 Fisher's exact test

As a second example out of many statistical significance test developed in mathematical statistics we mention *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 in general is a $m \times n$ matrix M where all possible outcomes of one variable x enter the columns in one row and distribution of outcomes of the second variable y is contained in the columns for a given row. 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

$$\begin{aligned} \text{probability mass function} \quad & f_{\mu,\nu}(k) = \frac{\binom{\mu}{k}\binom{N-\mu}{\nu-k}}{\binom{N}{\nu}} \ , \\ \text{cumulative density function} \quad & F_{\mu,\nu}(k) = \sum_{i=0}^{k} \frac{\binom{\mu}{k}\binom{N-\mu}{\nu-k}}{\binom{N}{\nu}} \ , \end{aligned} \tag{2.17}$$

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 probability of the general 2×2 contingency table is

$$p \ = \ \frac{\binom{a+b}{a}\binom{c+d}{c}}{\binom{N}{a+c}} \ = \ \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{a!\,b!\,c!\,d!\,N!} \ , \tag{2.18}$$

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 calculation yields $p \approx 0.10$ which is above the critical value $0.001 \leq \alpha \leq 0.05$ and $p > \alpha$ confirms the rejection of the assumption that men are more likely to be left-handed for these data.

3. Fisher-Wright debate and fitness landscapes

Although Ronald Fisher, J.B.S. Haldane, and Sewall Wright were united in their search for compatibility of Mendelian genetics and Darwinian selection, they differed strongly in their view on the existence and nature of a universal mechanism of evolution. In particular Fisher and Wright were engaged for more than thirty years in a heavy debate and each of both was more or less convinced that he had the solution to the problem [266]. No end of the debate occurred and no end was insight until Fisher's death in 1962. Interestingly the debate got a revival in 1997 when Jerry Coyne, Nick Barton and Michael Turelli [41] claimed that Fisher had the right theory and Wright's model is of minor importance if not dispensable at all. Inspired by this one-sided point of view Michael Wade and Charles Goodnight gave an answer in the same journal [297] wherein they argued that Fisher's theory cannot be applied to a variety of relevant phenomena in population genetics and is far away from being fully general. Accordingly, there is plenty of room for other theoretical approaches Wright's model being the most prominent one at the current state of knowledge [297]. In two follow-up papers [42, 120] the Fisher-Wright debate has been reignited and began to interest philosophers: Robert Skipper classified this scientific contest as a *relative significance controversy* and presented a proposal for a solution in the future that will be discussed in section 3.2. First, however, we shall present a new interpretation of Fisher's fundamental theorem that tries to rescue generality and relevance of this evolutionary optimization principle.

3.1 Fisher's fundamental theorem revisited

In section 2.5 Fisher's differential equation combining recombination and selection has been presented and analyzed. It describes the evolution of the allele distribution at a single locus. The variables are the normalized allele

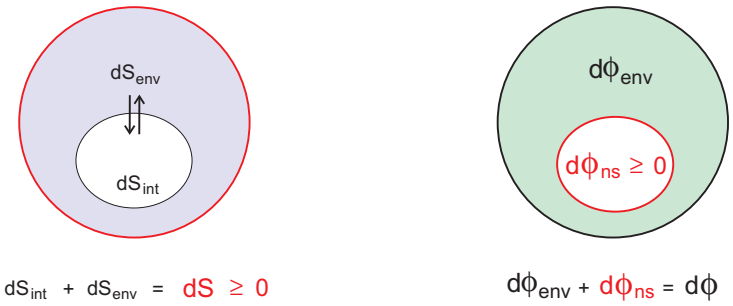


Figure 3.1: **Comparison of thermodynamical entropy and mean fitness.** The extremum principle of the entropy in thermodynamics applies to isolated systems, which are systems that sustain neither exchange of energy nor exchange of matter with their environment, and it is of the form $dS/dt \leq 0$ where S is the total entropy of the system (lhs of the figure). Isolated systems may house open systems that exchange energy and/or matter with their environment being part of the isolated system. The open system (white circle in the sketch) by itself does not fulfil the second law criterium because entropy can be exported to or imported from the environment: No maximum principle holds for S_{int}. Fisher's fundamental theorem is sketched on the rhs. The change in the mean population fitness is partitioned into two contributions, $\mathrm{d}\phi = \mathrm{d}\phi_{\text{ns}} + \mathrm{d}\phi_{\text{env}}$, out of which only one, $\mathrm{d}\phi_{\text{ns}}$ fulfils the maximum principle. Color code: The ranges of validity of the unidirectionality principle are indicated by red lines.

frequencies, $\sum_{i=1}^{n} x_i = 1$, the concentrations of the diploid genotypes are fully determined by the assumption of Hardy-Weinberg equilibrium. In essence, this is the *gene's eye view* as it has been popularized by Richard Dawkins [46]. The quantity that is nondecreasing and hence optimized is the mean reproduction rate of the allele distribution at the locus, $\phi(t) = \sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik} x_i x_k$. This function $\phi(t)$ obeying a *directionality principle* (Equ. (2.11)) was sometimes considered as an off-equilibrium equivalent to the entropy in equilibrium thermodynamics ([48]; for a more recent review see [49]), which according to the second law fulfils universal unidirectionality in isolated systems, $dS/dt \geq 0$. In Fig. 3.1 we present a sketch of the optimization principles in thermodynamics and in evolution. The major difference between both cases of unidirectionality concerns the range of validity: The thermodynamic principle holds globally, and need not be fulfilled in open subsystems whereas the

fundamental theorem is valid only for a subset of factors influencing mean fitness ϕ. This subset can be identified with natural selection. For reasons the will become clear within the next paragraphs the more recent and more elaborate interpretations of the fundamental theorem suggest that such an correspondence is not justified. It is, however, fair to say that Fisher himself stressed the limitation of the analogy [228, pp.346,347].

At first it is important to emphasize that Fisher's theorem as expressed by Equ. (2.11) is neither wrong not inexact. The (justified) critique concerns the very limited applicability. As said before the theorem does not apply in the two locus case when the two genes interact – a phenomenon that is called *epistasis*, there must be no linkage disequilibrium implying the necessity of random mating, and all other nonadditive genetic interactions must be zero. As we shall see later on, Fisher's *large population size theory* (L[P]ST)[1] several of the jeopardizing deviations become small and unimportant in the limit Fisher is considering [62, 87, 228, 236]. Secondly, as Ronald Fisher himself stressed several times, the (total) mean fitness of a population in nature can only fluctuate around zero because otherwise the population would either explode or collapse and die out.[2] How can this undeniable fact be reconciled with the Darwinian principle of optimizing fitness? The explanation of the contradiction The variation in the mean fitness is split into two different contributions: (i) the increase in mean fitness caused by natural selection, and (ii) the change in mean fitness caused by the environment where we define the environment as everything contributing to changes except natural selection:

$$\mathrm{d}\phi \;=\; \mathrm{d}\phi_{\mathrm{ns}} \,+\, \mathrm{d}\phi_{\mathrm{env}} \;, \tag{3.1}$$

and $\phi_{\mathrm{ns}} = \sum_{k=1}^{n} \sum_{i=1}^{n} a_{ik} x_i x_k$ is obeying the directionality principle. The notation of $\mathrm{d}\phi_{\mathrm{env}}$ being the change in mean fitness caused by the environment

[1] Fisher's theory is often abbreviated as LPST, sometimes as LST. We shall adopt here the shorter three letter version.

[2] Cases of such *explosions* are known but very rare. If, for example, a species is transferred into a new environment where other species predating on them are missing extremely rapid population growth occurs, which corresponds to a (temporarily) large positive mean fitness. Examples are the rabbits in Australia and several cases of insect proliferations causing major damage.

sounds weird, and makes sense only in the gene's eye view. Then, epistatic effects coming from other genes through interaction may *look* as environmental influence for the gene under consideration. The interpretation of other contributions to $\mathrm{d}\phi_{\mathrm{env}}$ is even more difficult. However, with this definition Fisher's fundamental theorem can be rescued also its range of applicability has shrunk to the small white area in Fig. 3.1 and definitely FTNS cannot be applied to the total $\mathrm{d}\phi$ as it is done in the conventional interpretation. There are still a number of problems with the subtler recent interpretation but most of them can be straightened out by careful argumentation [228]. A straightforward interpretation of the two views on Fisher's fundamental theorem concerns the nature of the time derivative:

$$\text{conventional view: } \frac{\mathrm{d}\phi}{\mathrm{dt}} \geq 0\,, \quad \text{new view: } \left.\frac{\partial\phi_{\mathrm{ns}}}{\partial t}\right|_{\mathrm{env}} \geq 0\,,$$

The conventional view was dealing with the total differential whereas the new view considers the partial differential at constant environment.

The important issue touched upon above can now find an answer. The gain in mean fitness of the allele population resulting from natural selection is compensated by the changes in the environment. Accordingly, we have $\mathrm{d}\phi_{\mathrm{env}} < 0$ or in other words for the gene the environment deteriorates to such an extent that the stationarity of the (eco)system is maintained. This compensation effect reminds of Leigh van Valen's *red queen hypothesis*:[3] *... and the Red Queen said to Alice: "In this place it takes all the running you can do, to keep in the same place. ..."*. This sentence is commonly used as a metaphor for the requirement of continuing development of an evolutionary system in order to maintain its (total) fitness relative to the (environmental) system it is coevolving with. Transferring the metaphor to Fisher's (peculiar) definition of environment this means for the allele population at a given locus: As its mean fitness increases by natural selection the environment deteriorates to about the same amount.

[3]The *Red Queen* is a fictional character in Lewis Carroll's fantasy novella, *Through the Looking-Glass*, which is often mistaken with the *Queen of the Heart* in Carroll's previous book *Alice's Adventures in Wonderland*.

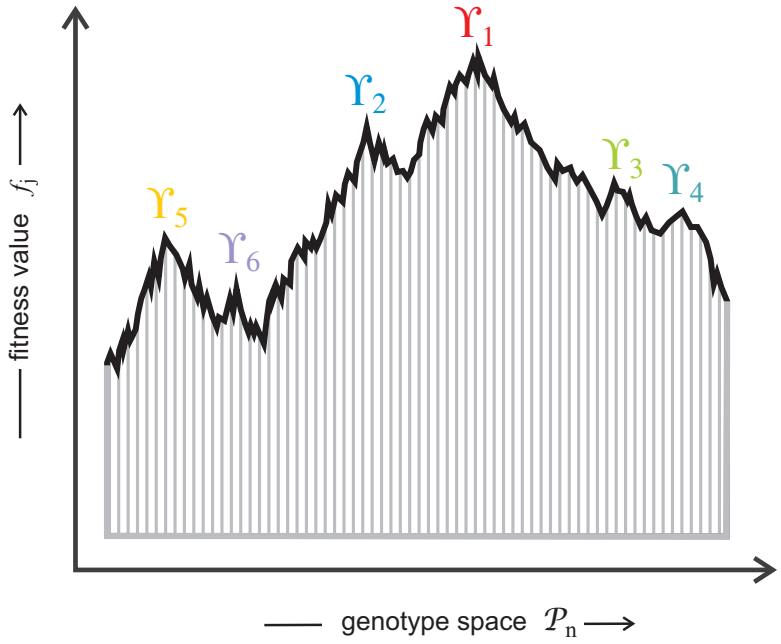


Figure 3.2: **Sewall Wright's fitness landscape.** The landscape has been introduced as a metaphor to illustrate evolution [319]. Populations or subpopulations of species are climbing on landscape with multiple peaks and optimize fitness in a non-descending or adaptive walk until they occupy local maxima. The fitness landscape is constructed through assigning a fitness value to every node of the support graph. Genotype space in Wright's original concept is recombination space and as such it is high-dimensional. In the simple sketch here the graph on which the landscape is plotted is a so-called *path graph* $\mathcal{P}_n$, which consists of n nodes on a straight line.

3.2 The Fisher-Wright controversy

Ronald Fisher's model of evolution, the large population size theory (LST), is based (i) on the assumption of large panmictic populations,[4] (ii) on mutation and natural selection as the major process driving evolutionary change, (iii) additive genetic effects and context independence of alleles, and (iv) refinement of existing adaptations in a stable and slowly changing environment

[4] *Panmixis* or *panmixia* means that there are no restrictions of any kind – be it genetic, physical or geographical – in mating. Mating partners are chosen fully at random.

as the ultimate driving force for evolution. Seen from this point of view the factors giving rise to deviations from the fundamental theorem are minor corrections of the global picture. For example, random drift plays a dominant role in small population, when genetic effects are predominantly additive and context independent, epistasis and pleiotropy are negligible.

Sewall Wright's model contrasts Fisher's view in many aspects. His model consists of three logical phases: (i) Random drift leads to semi-isolated subpopulations or demes within the global population, which are losing fitness because of accidental loss of fittest genotypes by the mechanisms of Mullers ratchet,[5] (ii) natural selection acts on complex genetic reaction networks and raises the mean fitness of subpopulations, (iii) interdemic selection raises the mean fitness of the global population. Eventually, environmental change shifts the adaptive peaks of mean fitness and drives the dynamics of evolution [228].

Depending on the species under consideration both evolutionary scenarios, the Fisher scenario and the Wright scenario, can be realized. None of the two models of evolution has internal inconsistencies that would allow for rejection. Thus the dispute between the two scholars is – what the philosophers call – a *relative significance controversy*. Both concepts are valid approaches that apply only to a limited subset of evolutionary scenarios and multiplicity of theoretical approaches is unavoidable at least at the current state of knowledge. In contrast to the view of most biologists and some philosophers (see, e.g., [266]) I think there is no need to believe that there will never be a uniform theory of evolution [249]. The unification, however, will not come on the phenomenological level of biology, it will be the result of a comprehensive theoretical biology that has its basis at the molecular level. The basis of this optimistic view comes from examples from physics: Electricity and magnetism were seen as largely unrelated phenomena unless the unifying theory of electromagnetism was born that found its elegant completion by James Clerk Maxwell who conceived the famous Maxwell equations.

[5]Hermann Joseph Muller considered a random drift process in a finite population. Stochasticity (see chapter 8) will cause a loss of the fittest genotype at some instant, then the fittest genotype of the rump population will be lost, and so on [90, 220, 221].

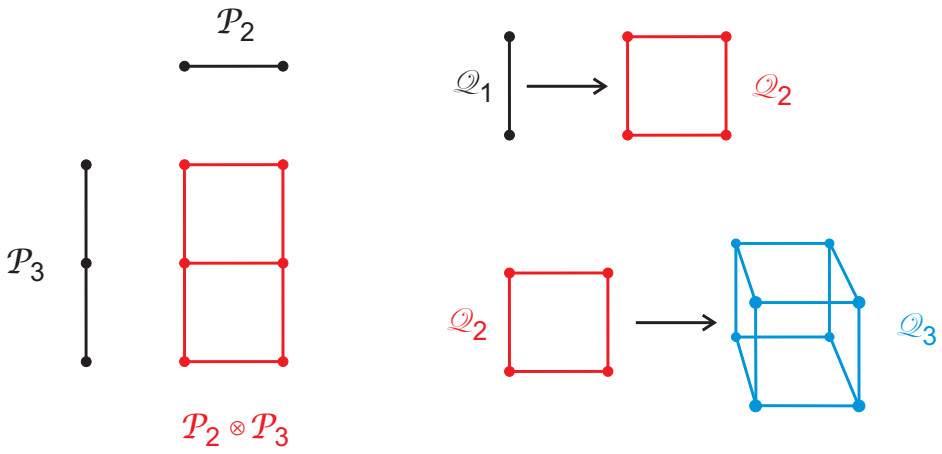


Figure 3.3: **Path graphs and binary sequence spaces**. The structure of binary sequence spaces $\mathcal{Q}_l$ follows straightforwardly from the graph Cartesian product of the path graph $\mathcal{P}_2$. The definition of the graph Cartesian product is shown on the lhs of the figure. The sketch on the rhs presents the construction of binary sequence spaces: $\mathcal{Q}_1 = \mathcal{P}_2$, $\mathcal{Q}_2 = \mathcal{P}_2 \otimes \mathcal{P}_2$, $\mathcal{Q}_3 = \mathcal{P}_2 \otimes \mathcal{Q}_2 = \mathcal{P}_2 \otimes \mathcal{P}_2 \otimes \mathcal{P}_2$, and so on, and $\mathcal{Q}_l = (\mathcal{P}_2)^l$.

3.3 Fitness landscapes on sequence spaces

Sewall Wright's original landscape metaphor was fitness plotted on recombination space as support. Recombination space is a discrete space with every possible allele combination or genome represented by a point. It is huge as Wright has already recognized: For the modest number of two alleles per gene and about 4000 genes for a bacterium like *Escherichia coli* there are $2^{4000} = 1.3 \times 10^{1204}$ combinations, and for the human genome with some 30 000 genes this number raises to $2^{30\,000} = 7.9 \times 10^{9030}$. These numbers are so far outside any imagination that one doesn't need to comment them. If we assume independent variation corresponding to the absence of linkage disequilibrium recombination space would have a dimension of several thousand but only two points in every direction – a bizarre object indeed but such a structure is typical for discrete spaces of combinatorial objects which are assembled from a set of building blocks. The structure of recombination

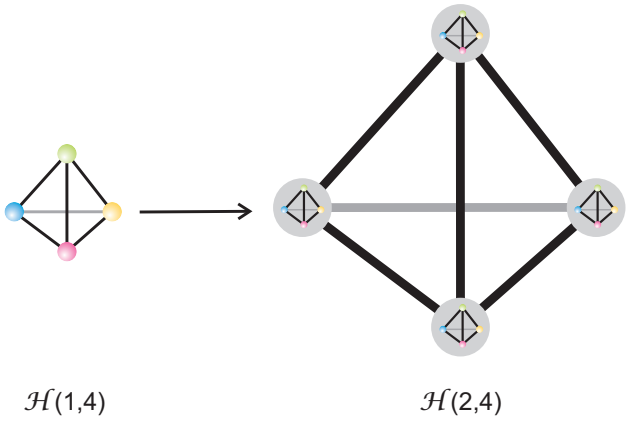


Figure 3.4: **Four letter sequence spaces**. The sequence space derived from the four letter alphabet (**AT**(**U**)**GC**; $\kappa = 4$) are the Hamming graphs $\mathcal{H}(l,4)$. The Hamming graph for a single nucleotide is the complete graph $\mathcal{H}(1,4) = \mathcal{K}_4$ (lhs) and for the 16 two letter sequences the space is $\mathcal{H}(2,4) = \mathcal{K}_4 \otimes \mathcal{K}_4$ (rhs). The general case, the space of sequences of chain length l, $\mathcal{H}(l,4)$ is the graph Cartesian product with l factors $\mathcal{K}_4$.

space has been studied in some detail [269, 272] but we dispense here from reviewing it, because we shall be mainly concerned with another discrete formal space, the *sequence space* here and in the forthcoming chapters.

Despite enormous progress in synthetic biology [137, 196] the possibilities to construct gene combinations at will are still very limited. For example, oscillatory gene networks [75] and genetic toggle switches [108] were engineered in bacteria, and various synthetic genetic regulatory elements were introduced into eukaryotic cells [11, 13], but engineered recombination is still not achievable at present. Engineering sequences with mutations at arbitrary positions has become routine thirty years ago already [96, 155, 184] and searching a space of sequences is much easier than searching recombination space therefore.

The idea of *sequence space* without the explicit notion has been used to order strings in informatics for quite some time (see, e.g., [133]) before the word has been coined for proteins [201] and nucleic acids [65]. Like recombi-

nation space sequence space is a discrete formal space where each sequence is represented by a point (see, for example, Fig. 4.7 where the sequence space of binary sequences of length l is used for illustration). The sequence space for binary sequences is a hypercube Q_l of dimension l where l is the sequence length of the string. The building principle of sequence spaces by means of the graph Cartesian product is illustrative and can be used for sequences over arbitrary alphabets and, in particular, also for the natural **AT**(**U**)**GC** alphabet. The Cartesian product of two graphs is illustrated in Fig. 3.3 by means of two path graphs:[6] The product graph, $\mathcal{P}^{(1)} \otimes \mathcal{P}^{(1)}$ is two-dimensional and has $\mathcal{P}^{(1)}$ on its horizontal and $\mathcal{P}^{(2)}$ on its vertical margin, respectively. There are many ways to visualize binary sequence spaces as hypercubes – one, the consecutive product of $\mathcal{P}_2$ graphs is illustrated in Fig. 3.3:

$$\mathcal{Q}_l \;=\; \mathcal{P}_2 \otimes \mathcal{P}_2 \otimes \ldots \otimes \mathcal{P}_2 \;= \left(\mathcal{P}_2\right)^l . \tag{3.2}$$

The advantage of the construction of sequence spaces a graph Cartesian products has the advantage of being generalizable. If we choose a complete graph $\mathcal{K}_\kappa$ as unit the consecutive Cartesian product yields the corresponding sequence space for sequences of chain length l:

$$\mathcal{Q}_l^{(\kappa)} \;=\; \mathcal{K}(l,\kappa) \;=\; \mathcal{K}_\kappa \otimes \mathcal{K}_\kappa \otimes \ldots \otimes \mathcal{K}_\kappa \;= \left(\mathcal{K}_\kappa\right)^l . \tag{3.3}$$

The most important case is the natural alphabet with $\kappa = 4$ (Fig. 3.4).

Both recombination and sequence spaces are characterized by high dimensionality and this makes it difficult to visualize distances. Considering, for example, the binary sequence space for strings of chain length $l = 10$ that contains $2^{10} = 1024$ sequences. Were sequence space a (one dimensional) path graph, the longest distance would be 1023. In two dimensions is would be 62 and on the hypercube $\mathcal{Q}_{10}^{(2)}$ it is shrunk to only 10. Both recombination and sequence space are metric spaces. The most natural metric for sequence spaces is the Hamming distance d_{H} (see section 4.3.1), for recombination spaces the construction an appropriate metric is somewhat more involved and can be done by using graph theory [272].

[6] A path graph $\mathcal{P}_n$ is a one-dimensional graph with n nodes. Two nodes at the ends have vertex degree one and all other $n-2$ nodes have vertex degree two.

A landscape is obtained through plotting some property on a graph as support (see Fig. 3.2). On a fitness landscape fitness values are assigned to the nodes of sequence space. The mathematical analysis of typical landscape properties, for example correlation functions, on discrete spaces require special techniques that are based on the use of Laplace operators on graphs [271]. Landscapes on supports that are combinatorial manifolds like the genomes obtained by recombination and/or mutation have been studied [238, 239] in some detail. Since the size of even the smallest sequence spaces is prohibitive for the empirical determination of a fitness landscape, simple models were used in particular in population genetics. The three most popular are (i) the additive fitness landscape, (ii) the multiplicative fitness landscape, and (iii) the single peak fitness landscape. Sequences are grouped around the fittest genotype. It is assumed that all sequences of a given mutant class have the same fitness. In the first case all mutations reduce the fitness of the fittest genotype by a constant increment θ, in the second case by a constant factor φ, and

$$\text{additive landscape}: \quad f_j = f_m - k\,\theta \quad \forall\, \mathbf{X}_j \in \Gamma_k \;, \tag{3.4a}$$

$$\text{multiplicative landscape}: \quad f_j = f_m\,\varphi^k \qquad \forall\, \mathbf{X}_j \in \Gamma_k \;, \quad \text{and} \tag{3.4b}$$

$$\text{single peak landscape}: \quad f_j = \begin{cases} f_0 \text{ if } j = m \\ f_n \text{ if } j \neq m \end{cases} \;, \tag{3.4c}$$

where Γ_k is the k-th mutant class of the fittest sequence $\mathbf{X}_m$:

$$\Gamma_k^{(m)}: \quad \mathbf{X}_i \in \Gamma_k^{(m)} \text{ iff } d_{\mathrm{H}}(\mathbf{X}_i, \mathbf{X}_m) = k \;. \tag{3.5}$$

Instead of the increment θ and the factor ϕ one might also use the lowest fitness value f_n for calibration and obtains $\theta = (f_0 - f_n)/l$ and $\phi = (f_n/f_0)^{1/l}$, respectively. As we shall later in a detailed discussion of the the properties of these different types of simple fitness landscapes (section 4.3.3) they are convenient for mathematical analysis but completely unrealistic.

Empirical knowledge and computer models of biopolymer – protein and nucleic acid – structures confirm to properties of realistic landscapes: (i) they are rugged and (ii) they are rich in neutral sequences. Rugged means that

nearby lying sequences – these are sequences of short Hamming distance – may have very different fitness values. This is easy to visualize since a single nucleotide exchange leading to an exchange of an amino acid that is essential for protein activity may render an enzyme functionless and in consequence the mutation might be lethal for its carrier. On the other hand, a single nucleotide exchange may have no effect at all and give rise to a neutral variant. What is true for a single nucleotide exchange already holds as well for multiple point mutations and other sequence changes Neutrality has been found early and predicted theoretically when the traces of evolution were discovered on the molecular level [173, 175]. The existence of a *molecular clock of evolution* is one important consequence of neutrality in evolution (for a review see [183]). One consequence of neutrality is random drift of populations through sequence space as described by Motoo Kimura's theory of neutral evolution [174] (see also section 10.3). Not all sequences are neutral, of course, but the neutral sequences form neutral networks in sequence space [238, 252] and these subsets of sequences set the stage for evolutionary dynamics.

The new techniques in molecular genetics provided access to empirical data of fitness landscapes for protein and nucleic acid evolution. Both basic features, ruggedness and neutrality were confirmed. Examples are protein fitness landscapes [140], landscapes derived from *in vitro* evolution of molecules [3], and a recent extensive study on the HIV-I fitness landscape in presence and absence of medication [181].

4. Mutation and selection

Gregor Mendel's concepts of segregation and independent assortment of genes during reproduction provides an excellent example of an approximate statistical law that has been modified by molecular insight but remained correct as the proper reference for the limit of infinitely frequent recombination. In case of mutation such an ideal reference does not exist and indeed this idea of change in the hereditary material has a fascinating history [34].

In this chapter we shall consider the consequences of mutation for evolutionary dynamics. We start by giving a brief historical account on the sharpening of the originally diffuse notion of mutation and then introduce chemical kinetics by means of a flow reactor that is useful for both, modeling and experimental studies. The kinetic model of a chemical reaction network based on ODEs introduced by Manfred Eigen in which correct replication and mutation are parallel reaction channels [65, 68, 70–72] replaces the *deus ex machina* appearance of mutants by. The formulation of the kinetic equations for overall replication,[1] that is without dwelling into molecular details is straightforward. Two major results of the mathematical analysis of these reaction networks are: (i) The formation of well defined stationary states called *quasispecies* and (ii) the existence of a maximal mutation rate characterized as *error threshold* above which no stationary states exist. This chemical reaction model of evolution has been directly applied to evolution of RNA molecules in the test-tube (see e.g. [14]), viroids and viruses, and bacteria as long as recombination plays very little or no role.

Reaction kinetics of RNA replication, for example, is a multi-step process that has been investigated at molecular resolution by Christof Biebricher [17–19] and the established mechanism can be incorporated into the model of

[1] Here we use *replication* when we have the molecular mechanism or some over-all kinetics of it in mind, and characterize the process of multiplication as *reproduction* when we don't refer to the molecular level or when the mechanisms is completely unknown.

evolution at any level of detail. In this case the extension of the overall kinetic description of evolution to a comprehensive model that treats replication and mutation at full molecular detail is possible and will be presented. Quasispecies theory has been applied to virus populations by Esteban Domingo and turned out to be the proper concept for a description of virus infection and virus evolution [51, 54, 191]. The error threshold phenomenon provides a proper frame for the development new antiviral strategies [56]. A dispute arose whether or not viral quasispecies pass the error threshold before the die out [30, 31, 283] and we shall present a brief overview of models for lethal mutagenesis.

For bacteria or higher organism the kinetic model of evolution can be applied in the same spirit as the ODE models in population genetics but the full details of reproduction are still in the dark. Nevertheless much progress has been made within the last ten years The organisms that come closest to full understanding in the spirit of systems biology are small parasitic bacteria of the species *Mycoplasma*. Examples are the works of the Spanish group of Luis Serrano's [128, 182, 322] who have almost completed the full biochemistry of an entire small bacterial cell. About 75 % of the approximately 6 000 genes of a small eukaryote, the yeast *Saccharomyces cerevisiae*, where studied with respect to their function and a quantitative genetic interaction profile has been established [38]. Functional cross-connections between all bioprocesses were derived from the global network and mapped onto a cellular wiring diagram of pleiotropy. These investigations in essence tell us two things: (i) The complexity of cellular biology at the molecular level is enormous, and (ii) the modern techniques of genomics, proteomics, transcriptomics, metabolomics, and so on, together with conventional biochemistry and biochemical kinetics can be successfully applied to decipher all processes in an entire cell at molecular resolution, and the data are usable for mathematical modeling and computer simulation at all levels of complexity.

4.1 The historical route to mutation

Systematic mutation studies were initiated in Thomas Hunt Morgan's laboratory. He began his seminal work with the fruit fly *Drosophila melanogaster* as animal model around 1910 and was awarded the Nobel Prize 1933 for Physiology for his extensive and path-breaking genetic studies. Later *Drosophila melanogaster*[2] became the standard animal for genetic studies mainly because of its enormous fertility and short (mean) generation time of only 7 to 60 days depending on subspecies on nutrition condition. Morgan's contributions were frequently dealing with genes on sexual chromosomes where the mechanism of gene action can be inferred and discussed more easily because every species provides two different chromosomal types, male and female. In the early days of genetics the term mutation was used for all inheritable changes of phenotypes, no matter whether the cause was a change in the gene or a rearrangement of the chromosomal structure of the cell. Known, of course, was the distinction between mutation along the germ line and *somatic mutation* that was not inheritable. The American biologist Hermann Muller argued that the term *mutation* should be restricted to *variation due to change in the individual gene* [218] and chromosomal changes should be considered separately since they have a different origin. Examples are polyploidy, gains or losses of entire sets of chromosomes, aneuploidy, gains or losses of single chromosome in otherwise diploid organisms, and structural rearrangements of chromosomes. The notion of *point mutation* had been created for spontaneous changes in phenotypic properties determined by a single locus, was not clearly distinguished from the mutation, and got his meaning only with the discovery of DNA as genetic material.

The idea to increase the natural mutation rate by radiation, mainly X-rays, in order to produce a diversity of variants was already pursued by the Morgan group and others but little progress was made. The early attempts to use radiation failed for several reasons, one of them was the lack of reliable dosimeters and the application of wrong doses – low dose for long times.

[2]Due to a more recent classification of the *Drosophilidae* the famous model animal belongs now to the subgenus *Sophophora* and the official name is *Sophophora melanogaster*.

Muller and others found that the occurrence of point mutations increases linearly with time whereas chromosomal rearrangements showed an exponential time law. By applying high doses Muller was able to increase the mutation rate by a factor of 150. The technique and the results of Muller's systematic studies on X-ray induced mutation [219] were also of great importance for agriculture where human selection of desired variants initiated a green revolution. Useful mutants were nevertheless rare as the majority of radiation induced variants that showed a phenotype were lethal. Muller was awarded the Nobel Prize for Physiology in 1946 for *the discovery of the production of mutations by means of X-ray irradiation*. Muller also tried to distinguish effects from primary radiation impacts, which are mainly identified with double strand breaks or point mutations originating from damage repair, and secondary effects that result from chemical mutagenesis caused by the products of radiation impact on water. Muller's view was that the secondary effects dominate and candidates for chemical mutagenesis were among others hydrogen peroxide – already close to the most potent agent produced by irradiation of water: the hydroxy radical. Many other chemicals were found to produce or promote mutation, but time was not yet ripe for chemical models of mutation because the molecular nature of the genetic material was still (at least to a large extent) in the dark. In particular the clear separation of variants producing processes into mutation, recombination, and major chromosomal rearrangements was impossible without a knowledge of the molecular structure and function of DNA.

The next major breakthrough in understanding mutation came from biochemistry. George Wells Beadle and Edward Lawrie Tatum found that mutations (can) change the metabolism of the affected mutation carrier [12]. Mutations change genes and the changes genes give rise to modification in an enzyme that is active in cellular metabolism: The *one gene - one enzyme* hypothesis of biology was born. Together with Joshua Lederberg Beadle and Tatum received the Nobel Prize 1958 in Physiology. Although time was already overripe for a major step forward in the understanding of the mechanisms of reproduction and inheritance, progress in the knowledge about nucleic acids was relatively slow [167]. It was Watson and Crick's model of

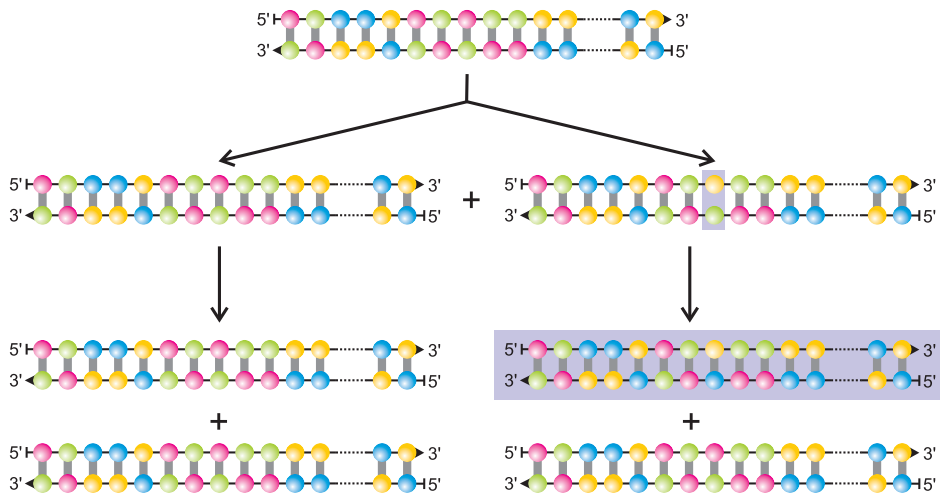


Figure 4.1: **Point mutation in DNA replication**. A misincorporation of a nucleotide during DNA replication leads to unpaired nucleotides at one position of the DNA double helix (second line, rhs: highlighted). The next replication step results in one correct copy and one mutant (third line, rhs: highlighted). The nucleotides are sketched as beads. Watson-Crick base pairs (**A**=**T** and **G**≡**C**) are shown a thick gray lines connecting two beads on opposite strands. Color code: **A** green, **T** magenta, **G** blue, and **C** yellow.

DNA structure that changed the situation completely, and the reason for this new insight is hidden in the famous sentence at the end of the letter to *Nature*: *"It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material"*. As a matter of fact the DNA structure suggested at the same time a mechanism for the origin and inheritance of point mutations (Fig. 4.1). On the other hand the consequences of mutations for fitness and evolution were completely in the dark and still are unknown in many aspects at present.

Many different classes of mutations and major DNA rearrangements were observed so far. The three most important classes are sketched in Fig. 4.2. Apart from point mutations deletions and insertions occur rather often. Deletions represent losses of a wide range of sizes, from individual nucleotides to whole genes or chromosomes. Insertions commonly are duplication of DNA stretches and they comprise an even wider range of seizes from single nu-

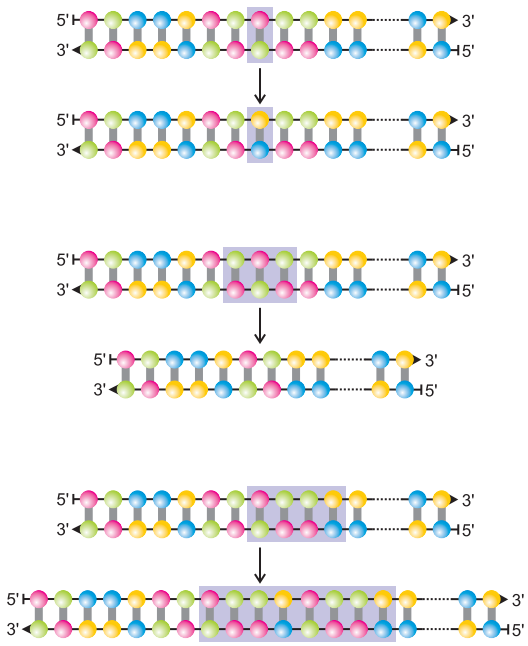


Figure 4.2: **Classes of mutations**. The sketch shows the three most common classes of DNA mutations: (i) point mutations (top, see Fig. 4.1), (ii) deletions (middle), and (iii) insertions (bottom). Deletion and insertions can be of any size. Examples for large scale events are gene loss and gene duplication. Eventually, whole genome duplication events have occurred in evolution. Color code: **A** green, **U** magenta, **G** blue, and **C** yellow.

cleotide duplications to gene duplication and duplications of entire genomes. Gene duplication has been postulated as a major driving force for evolution by the Japanese Biologist Susumo Ohno [226]. Over the years many instances of genome duplications that have occurred in the past were found [289]. In particular, the role of genome duplications in developmental were studied in detail and the evolution of the *Hox* genes was found to present excellent examples for such events [59, 106, 185]. An ancient genome duplication event in the yeast *Saccharomyces cerevisiae* is especially well documented [168].

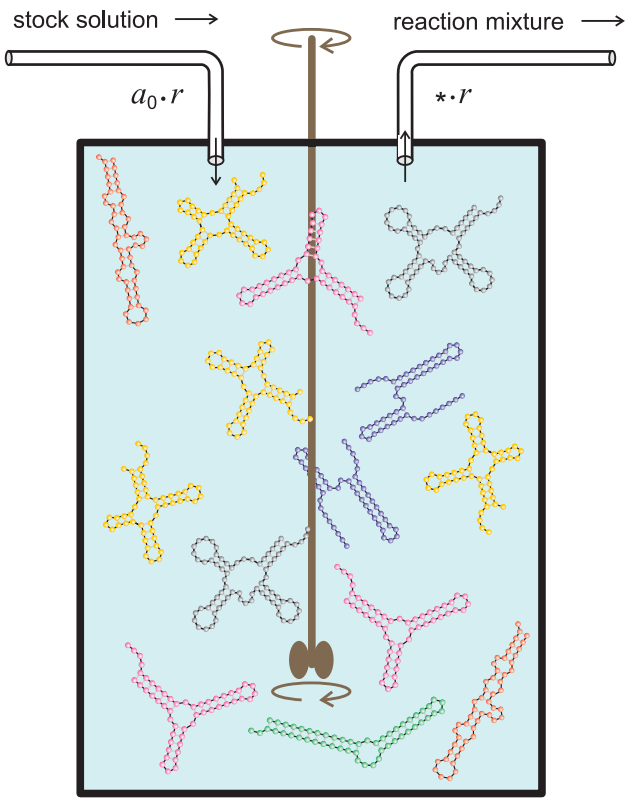


Figure 4.3: **The continuously stirred tank reactor (CSTR)**. The CSTR is a device for simulating deterministic and stochastic kinetics in an open system. A stock solution containing all materials for RNA replication ($[\mathbf{A}] = a_0$) including an RNA polymerase flows continuously at a flow rate r into a well stirred tank reactor (CSTR) and an equal volume containing a fraction of the reaction mixture ($[\star] = \{a, b, c_i\}$) leaves the reactor. A population of RNA molecules in the reactor ($\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ present in the numbers $N_1, N_2, \ldots, N_n$ with $N = \sum_{i=1}^{n} N_i$) fluctuates around a mean value, $N \pm \sqrt{N}$. RNA molecules replicate and mutate in the reactor, and the fastest replicators are selected. The RNA flow reactor has been used also as an appropriate model for computer simulations [99, 100, 156, 234], which allow for the application of other criteria for selection than fast replication. For example, fitness functions are defined that measure the distance to a predefined target structure and then the mean fitness of the population increases during the approach towards the target [101].

4.2 The flow reactor

The theoretical and experimental device of a flow reactor shown in Fig. 4.3 represents an open system that is perfectly suited for studies on evolutionary dynamics. The setup explains its name *continuously stirred tank reactor* (CSTR). A basic assumption of the theoretical model is instantaneous and complete mixture of the reactor content at any moment, and spatial homogeneity of the solution in the reaction volume allows for the usage of ordinary rather than partial differential equations for modeling reaction kinetics. Here we consider replication of RNA molecules as an example of an ensemble of competitive autocatalytic reactions in the spirit of the selection Equ. (1.9). The reaction mechanism is of the form

$$* \xrightarrow{a_0\, r} \mathbf{A}\ , \tag{4.1a}$$

$$\mathbf{A} + \mathbf{X}_j \xrightarrow{k_j} 2\,\mathbf{X}_j\,;\ j = 1, \ldots, n\ , \tag{4.1b}$$

$$\mathbf{X}_j \xrightarrow{k_j} \mathbf{B}\,;\ j = 1, \ldots, n\ , \tag{4.1c}$$

$$\mathbf{A} \xrightarrow{r} \varnothing\ , \tag{4.1d}$$

$$\mathbf{B} \xrightarrow{r} \varnothing\ , \quad \text{and} \tag{4.1e}$$

$$\mathbf{X}_j \xrightarrow{r} \varnothing\,;\ j = 1, \ldots, n\ . \tag{4.1f}$$

The symbol $\mathbf{A}$ stands for the material required in RNA synthesis and $\mathbf{B}$ are the degradation products of the RNA molecules.[3] The flow rate parameter r is the reciprocal mean residence time of a volume element in the reactor, $r = \tau_R^{-1}$. If the reactor has a volume V then a mean volume V has flown through it in the time interval $\Delta t = \tau_R^{-1}$. The influx of $\mathbf{A}$ into the reactor is a so-called zeroth order reaction, because the rate $\mathrm{d}\mathbf{A} = a_0 r\ \mathrm{dt}$ does not depend on concentrations in reaction step (4.1a). The autocatalytic second order reaction step (4.1b) is identical to the reproduction step in Equ. (1.8) only here the concentration of $\mathbf{A}$ is a variable $[\mathbf{A}] = a(t)$. The degradation reaction of $\mathbf{X}_j$ (4.1c) can be united with the outflux of $\mathbf{X}_j$ (4.1f) yielding a

[3] For the sake of simplicity we lump all building blocks $\mathbf{A}_i$ and all degradation products $\mathbf{B}_i$ $(i = 1, 2, \ldots)$ together in a single symbol.

rate parameter of $k_j + r$ unless **B** enters the system again via a chemical or photochemical recycling reaction $\mathbf{B} + \mathbf{D} \rightarrow \mathbf{A}$ or $\mathbf{B} + h\nu \rightarrow \mathbf{A}$. The reaction steps (4.1d-4.1f) eventually describe the outflux of the reactor content that compensates the increase in volume through influx of stock solution.

The mechanism (4.1) can be casted into a chemical master equation or modeled by stochastic processes. Here we shall assume large population sizes and use ODEs for modeling. The kinetic equations, with $[A] = a$ and $[X_j] = c_j$, are of the form

$$\begin{aligned} \frac{\mathrm{d}a}{\mathrm{dt}} &= -a \sum_{j=1}^{n} k_j c_j + r\,(a_0 - a)\ , \\ \frac{\mathrm{d}c_j}{\mathrm{dt}} &= c_j \left(a\,k_j - r\right)\ ;\ j = 1, \ldots, n\ , \end{aligned} \tag{4.2}$$

and can be readily analyzed with respect to long time behavior. For this goal we introduce a total concentration of replicators, $c = \sum_{i=1}^{n} c_i$, that fulfils

$$\frac{\mathrm{d}c}{\mathrm{dt}} = c\,(a\,\phi - r) \quad \text{with} \quad \phi(t) = \frac{1}{c} \sum_{i=1}^{n} k_i\, c_i \tag{4.3}$$

being the mean replication rate parameter. The total concentration of material in the flow reactor, $C = a + c$, fulfils the kinetic equation

$$\frac{\mathrm{d}C}{\mathrm{dt}} = \frac{\mathrm{d}a}{\mathrm{dt}} + \frac{\mathrm{d}c}{\mathrm{dt}} = r\,(a_0 - C)\ .$$

From $dC/dt = 0$ follows the stationarity condition $\bar{C} = \bar{a} + \bar{c} = a_0$, and two more conditions derived from $da/dt = 0$ and $dc/dt = 0$ yield:

(i) $P^{(1)}$: $\bar{c}^{(1)} = 0$ and $\bar{a}^{(1)} = a_0$,

(ii) $P^{(2)}$: $\bar{c}^{(2)} = a_0 - r/\bar{\phi}$ and $\bar{a}^{(2)} = r/\bar{\phi}$.

Thus, we have a state of extinction, $P^{(1)}$, with $\bar{c} = 0 \Rightarrow \bar{c}_j = 0\ (j = 1, \ldots, n)$, and n active states P_j belonging to $P^{(2)}$), with different combinations of the individual concentrations $\bar{c}_j$. For these states the conditions $dc_j/dt = 0$ are fulfilled by n different solutions

$$\begin{aligned} &\bar{c}_i = 0 \ \text{ for all } \ i \neq j\,;\ i = 1, \ldots, n\,,\ \bar{c}_j = a_0 - \frac{r}{k_j} \ \text{ and} \\ &\bar{a} = \frac{r}{k_j} \ \text{ for } \ j = 1, \ldots, n \ \text{ and } \ r < k_j\, a_0 \end{aligned} \tag{4.4}$$

where for a physically meaningful stationary point P_j all concentrations have to be nonnegative.

In order to derive the stability of the different stationary states we calculate the Jacobian matrix

$$\mathrm{A} \;=\; \begin{pmatrix} -\sum_{i=1}^{n} k_i c_i - r & -k_1 a & -k_2 a & \dots & -k_n a \\ k_1 c_1 & k_1 a - r & 0 & \dots & 0 \\ k_2 c_2 & 0 & k_2 a - r & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ k_n c_n & 0 & 0 & \dots & k_n a - r \end{pmatrix} .$$

At the stationary point P_j with $\bar{c}_j(P_j)$ and $\bar{a}(P_j)$ as given in Equ. (4.4) we obtain

$$\mathrm{A} \;=\; \begin{pmatrix} -k_j a_0 & -\frac{k_1}{k_j} r & \dots & -r & \dots & -\frac{k_n}{k_j} r \\ 0 & \left(\frac{k_1}{k_j} - 1\right) r & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ k_j a_0 - r & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & \left(\frac{k_n}{k_j} - 1\right) r \end{pmatrix} .$$

The corresponding eigenvalue problem can be solved analytically and yields for P_j:

$$\begin{aligned} \lambda_0 &= -r\,, \\ \lambda_1 &= \frac{k_1 - k_j}{k_j}\, r\,, \\ &\vdots \\ \lambda_j &= -k_j\, a_0 \,+\, r\,, \\ &\vdots \\ \lambda_n &= \frac{k_n - k_j}{k_j}\, r\,, \end{aligned} \tag{4.5}$$

A stationary state is stable if and only if all eigenvalues are negative. Considering Equ. (4.5) λ_0 is always negative, λ_j is negative provided $r < k_j\, a_0$,

and the sign of all other eigenvalues $\lambda_i\,(i \neq j)$ is given by the sign of the difference $k_i - k_j$. In the non-degenerate case – the parameters k_i are all different – the only stable stationary state P_m with $j \equiv m$ is defined by

$$k_m \;=\; \max\{k_1, k_2, \ldots, k_n\}\;.$$

From $\lambda_m = -k_m a_0 + r$ follows that P_m is also the state that has only positive concentrations up to the largest flow rate, $r = k_m\, a_0$, see (4.4).

The analysis of (competitive) replication in the flow reactor provides three results: (i) Flow rates above a critical limit lead to extinction, (ii) within a given ensemble of replicating molecules selection in the Darwinian sense of survival of the fittest takes place, i.e. the molecular species with the largest replication rate constant $f_m = \max\{f_j\}$, and (iii) the mean replication rate parameter of a population is the quantity that is optimized.

A modified flow reactor with automatic control of the flow rate facilitates the analysis of the kinetic differential equations. The flow rate $r(t)$ is regulated such that the total concentrations of replicators is constant, $c = c_0$:[4]

$$\sum_{i=1}^{n}\frac{dc_i}{dt} \;=\; 0 \;=\; \sum_{i=1}^{n} c_i(a\,k_i \,-\, r) \;=\; c_0\,(a\cdot\phi \,-\, r)\,,\quad \text{and}$$

$$r \;=\; a\cdot\phi \quad . \tag{4.6}$$

Equ. (4.6) is readily interpreted: In order to keep the concentration of replicators constant, the flow rate has to be raised when either the concentration of **A** or the mean replication rate of the population, ϕ, increases. The conservation relation $c = c_0$ reduces the number of independent variables from $n+1$ to n: a and $n-1$ concentration variables c_j. The kinetic equations can be written now

$$\frac{\mathrm{d}a}{\mathrm{dt}} \;=\; a\,\phi\,(a_0 - c_0 - a) \;\;\text{and}$$

$$\frac{\mathrm{d}c_j}{\mathrm{dt}} \;=\; a\,c_j\;(k_j \,-\, \phi)\;,\;\; j = 1, 2, \ldots, n-1\;.$$

[4]Reactors called *cellstat* or *turbidostat* serve this goal. Such devices are used, for example, in microbiology to maintain constant concentrations of bacteria [29, 153, 154, 197]. The concentration of particles is monitored and regulated by parameters like optical turbidity or dielectric permittivity.

The remaining concentration and the mean replication rate parameter are defined by

$$c_n \;=\; c_0 - \sum_{i=1}^{n-1} c_i \;\;\text{and}\;\; \phi \;=\; \frac{1}{c_0}\left(k_n c_0 + \sum_{i=1}^{n-1}(k_i - k_n)c_i\right) \;.$$

Multiplication of the time axis by the factor $a(t) > 0$, $d\tau = a(t)\,dt$, yields:

$$\begin{aligned} \frac{\mathrm{d}c_j}{\mathrm{d}\tau} \;&=\; c_j\,(k_j \,-\, \phi)\;,\;\; j = 1, 2, \ldots, n-1\,, \;\;\text{and} \\ \frac{\mathrm{d}a}{\mathrm{d}\tau} \;&=\; (a_0 - c_0 - a)\,\phi\,, \;\;\text{with}\;\; \frac{\mathrm{d}\phi}{\mathrm{d}\tau} \;=\; \mathrm{var}\{k\}\,. \end{aligned} \tag{4.7}$$

Equ. (4.7) is self-contained in the sense that it does not require information on $a(t)$ to be solved, although $a(t)$ is contained in the transformed time axis τ. The coupled differential equation for $a(t)$ on the other hand requires explicit knowledge of the variables $c_j(t)$, which appear in $\phi(t)$. Equ. (4.7) is also formally identical with Equ. (1.14), which has been derived under the assumption of constant total concentration, $\sum_{i=1}^{n} c_i = c_0$ and constant concentration of $\mathbf{A}$: $[\mathbf{A}] = a_0$, which corresponds to a large reservoir of building blocks. Then, a_0 can be incorporated into the rate parameters, $f_j = k_j \cdot a_0$, obtaining thereby the analog of Equ. (1.14):

$$\frac{\mathrm{d}c_j}{\mathrm{dt}} \;=\; c_j\,(f_j \,-\, \phi)\;;\;\; j = 1, 2, \ldots, n-1\,. \tag{4.7'}$$

This condition simplifying the analysis of the kinetic equations is called *constant organization* in the kinetic theory of molecular evolution [71] and used also in population genetics. As far as the dimensions of the variables, the rate parameters and the mean replication rate is concerned there is a subtle difference: $[k] = [\phi] = [\mathrm{time}^{-1} \cdot \mathrm{concentration}^{-1}]$ and $[\tau] = [\mathrm{time} \cdot \mathrm{concentration}]$ in Equ. (4.7), whereas in the equation above $[f] = [\phi] = [\mathrm{time}^{-1}]$ and $[t] = [\mathrm{time}]$. The two ODEs (4.7) and (4.7') are a nice example of two dynamical systems with identical trajectories and different solution curves because of different time axes.

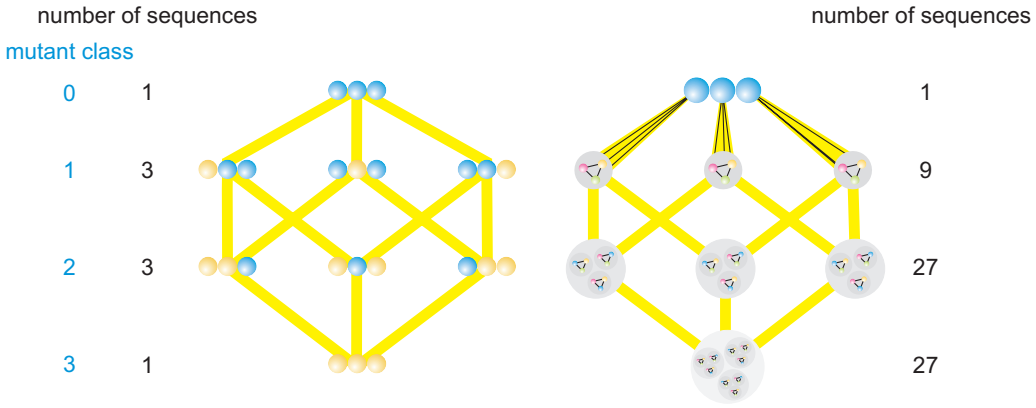


Figure 4.4: **Consecutive mutations**. Consecutive point mutations propagate into sequence space and are properly group in mutant classes Γ_k where k is the Hamming distance from the reference sequence in Γ_0. For binary sequences ($\kappa = 2$) the numbers of sequences in mutant class k follows the binomial distribution: $|\Gamma_k| = \binom{l}{k}$, and in the general case we have $|\Gamma_k| = (\kappa-1)^k\binom{l}{k}$. The figure sketches the sequence spaces of binary (lhs) and natural (**AUGC**) sequences (rhs) of chain length $l = 3$.

4.3 Replication and mutation

The interplay of replication, mutation, and selection is a core issue of Darwinian evolution, which could not be properly approach before knowledge on structures and functions of the molecules involved in the process became available. In particular, the accessibility of mutants requires knowledge on the mechanism of mutation and the internal structure of a mutation space. In section 3.3 we discussed the sequence space as the mutation compatible ordering principle of sequences and considered it as the appropriate support for fitness landscapes. Mathematically the internal structure of sequence space is given by the properties of the Hamming graph $\mathcal{H}(l, \kappa)$ where l is the chain length of the sequence and κ the size of the nucleobase alphabet. Evolution proceeding via consecutive mutations is confined by the properties of sequence space as sketched in Fig. 4.4 for the paths determined by point mutations. The size of the nucleobase alphabet clearly determines the diversity

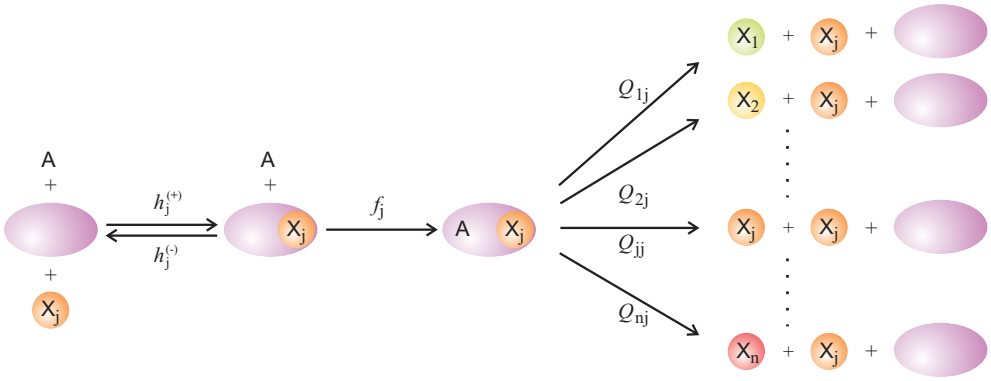


Figure 4.5: **A molecular view of replication and mutation**. The replication device $\mathbf{E}$, commonly a replicase molecule or a multi-enzyme complex (violet) binds the template DNA or RNA molecule ($\mathbf{X}_j$, orange) in order to form a replication complex $\mathbf{E}\cdot\mathbf{X}_j$ with a binding constant $H_j = h_{+j}[\mathbf{E}][\mathbf{X}_j] \,/\, h_{-j}[\mathbf{E}\cdot\mathbf{X}_j]$ and replicates with a rate parameter f_j. During the template copying process reaction channels leading to mutations are opened through replication errors. The reaction leads to a correct copy with frequency Q_{jj} and to a mutant $\mathbf{X}_k$ with frequency Q_{kj} commonly with $Q_{jj} \gg Q_{kj}\,\forall\, k \neq j$. Stoichiometry of replication requires $\sum_{i=1}^{n} Q_{ij} = 1$, since the product has to be either correct or incorrect. The reaction is terminated by full dissociation of the replication complex. The sum of all activated monomers is denoted by $\mathbf{A}$.

of mutations. For a binary sequence of length l the number of mutants with k errors being the number of sequences in the mutant class Γ_k is simply given by the binomial distribution: $|\Gamma_k| = \binom{l}{k}$. The formula is easily generalized to an alphabet with κ letters: $|\Gamma_k(\kappa;l)| = (\kappa-1)^l\binom{l}{k}$. The change in population structure due to mutation is determined by the branching structure of the *mutation tree*. Since mutant diversity increases fast with increasing number of consecutive mutants being tantamount to the Hamming distance d_{H} from the reference sequence the question arises, how large is the fraction of mutants that can be tolerated by intact inheritance. After introducing a mathematical model for mutation based evolution we shall return to the problem of error tolerance in consecutive replication.

4.3.1 Mutation-selection equation

As follows from the logic of mutation shown in Fig. 4.1 correct replication and mutation are parallel chemical reactions, which are initiated in the same way. A sketch of such a mechanism is shown in Fig. 4.5: A replication device – a single single-strand DNA replicating enzyme in the polymerase chain reaction (PCR), an RNA-specific RNA polymerase in most cases of viral RNA replication or a large multi-enzyme complex in DNA double-strand replication – binds the template, replication is initiated and correct replication and mutations represent different replication channels, since replication errors occur on the fly. When replication is completed the complex between template, product, and replication device dissociates. We distinguish direct replication, $\mathbf{X} \rightarrow 2\mathbf{X}$, and complementary replication, $\mathbf{X}^{(+)} \rightarrow \mathbf{X}^{(-)} + \mathbf{X}^{(+)}$ and $\mathbf{X}^{(-)} \rightarrow \mathbf{X}^{(+)} + \mathbf{X}^{(-)}$, where the former is typical for DNA replication of all organisms and the latter occurs commonly with RNA single-strand viruses. Here, we start by considering direct application first and shall show later that complementary replication under almost all reasonable conditions can be well approximated by a single overall direct replication step.

In order to introduce mutations into selection dynamics Manfred Eigen [65] conceived a kinetic model based on overall stoichiometric equations, which handle correct replication and mutation of an asexually reproducing species as parallel reactions (Fig. 4.5)

$$(\mathbf{A}) + \mathbf{X}_i \xrightarrow{Q_{ji} f_i} \mathbf{X}_j + \mathbf{X}_i \,;\; i, j = 1, \ldots, n \;. \tag{4.8}$$

Since $\mathbf{A}$ is assumed to be present in excess it is a constant and contained as a factor in the rate parameter $f_j = a_0\, k_j$

In normalized coordinates (4.8) corresponds to a differential equation of the form

$$\frac{\mathrm{d}x_j}{\mathrm{dt}} \;=\; \sum_{i=1}^{n} Q_{ji} f_i\, x_i \;-\; x_j\, \phi(t) \,;\; j = 1, \ldots, n \;;\; \sum_{i=1}^{n} x_i = 1 \;. \tag{4.9}$$

The finite size constraint $\phi(t) = \sum_{i=1}^{n} f_i x_i$ is precisely the same as in the mutation-free case (1.14), and the integrating factor transformation [329,

p. 322ff.] can be used to solve the differential equation [165, 285]:

$$z_j(t) \;=\; x_j(t)\,\exp\left(\int_0^t \phi(\tau)\mathrm{d}\tau\right) \quad \text{with} \quad \exp\left(\int_0^t \phi(\tau)d\tau\right) \;=\; \sum_{i=1}^{n} z_i(t)\;,$$

Insertion into Equ. (4.9) yields

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left(z_j(t)\,e^{-\int_0^t \phi(\tau)d\tau}\right) \;&=\; \frac{\mathrm{d}z_j}{\mathrm{d}t}\,e^{-\int_0^t \phi(\tau)d\tau} \;+\; z_j(t)\,\frac{\mathrm{d}}{\mathrm{d}t}\,e^{-\int_0^t \phi(\tau)d\tau} \;= \\
&=\; \frac{\mathrm{d}z_j}{\mathrm{d}t}\,e^{-\int_0^t \phi(\tau)d\tau} \;-\; z_j(t)\,e^{-\int_0^t \phi(\tau)d\tau}\,\phi(t) \;= \\
&=\; \left(\sum_{i=1}^{n} Q_{ji}f_i\,z_i(t) \;-\; z_j(t)\,\phi(t)\right) e^{-\int_0^t \phi(\tau)d\tau}\;, \\
\frac{\mathrm{d}z_j}{\mathrm{d}t} \;&=\; Q_{ji}f_i\,z_i(t) \quad \text{or} \quad \frac{\mathrm{d}\mathbf{z}}{\mathrm{d}t} \;=\; \mathrm{Q}\cdot\mathrm{F}\cdot\mathbf{z} \;=\; \mathrm{W}\cdot\mathbf{z}\;. \qquad (4.10)
\end{aligned}$$

The transformation turns the mildly nonlinear ODE (4.9) into a linear equation, which can be solved by standard mathematical techniques. All parameters are contained in the two $n \times n$ matrices, the *mutation matrix* Q, the *fitness matrix* F: The mutation frequencies are subsumed in the matrix $\mathrm{Q} = \{Q_{ji}\}$ with Q_{ji} being the probability that $\mathbf{X}_j$ is obtained as an error copy of $\mathbf{X}_i$. The fitness values are the elements of a diagonal matrix $\mathrm{F} = \{F_{ij} = f_i\,\delta_{i,j}\}$ and the *value matrix* W finally, is the product of the two matrices: $\mathrm{W} = \mathrm{Q}\cdot\mathrm{F} = \{W_{ji} = Q_{ji}f_i\}$.

Provided matrix W is diagonalizable, which will always be the case if the mutation matrix Q and the fitness matrix F are based on real chemical reaction mechanisms, we can transform variables by means of the two $n \times n$ matrices $\mathrm{B} = \{b_{ij}\}$ and $\mathrm{B}^{-1} = \mathrm{H} = \{h_{ij}\}$ $(i, j = 1, \ldots, n)$,

$$\mathbf{z}(t) \;=\; \mathrm{B}\cdot\boldsymbol{\zeta}(t) \;\text{ and }\; \boldsymbol{\zeta}(t) \;=\; \mathrm{B}^{-1}\cdot\mathbf{z}(t) \;=\; \mathrm{H}\cdot\mathbf{z}(t)\;,$$

such that $\mathrm{B}^{-1}\cdot\mathrm{W}\cdot\mathrm{B} = \Lambda$ is diagonal and its elements, λ_k, are the eigenvalues of the matrix W. The right-hand eigenvectors of W are given by the columns of B, $\mathbf{b}_j = (b_{i,j};\, i = 1, \ldots, n)$, and the left-hand eigenvectors by the rows of $\mathrm{B}^{-1} = \mathrm{H}$, $\mathbf{h}_k = (h_{k,i};\, i = 1, \ldots, n)$, respectively. These eigenvectors are the *normal modes* of mutation-selection kinetics.

For strictly positive off-diagonal elements of W, implying the same for Q, which implies nothing more than every mutation $\mathbf{X}_i \rightarrow \mathbf{X}_j$ has to be possible although the probability of occurrence might be extremely small, Perron-Frobenius theorem holds (see, for example, [259] and below) and we are dealing with a non-degenerate largest eigenvalue λ_0,

$$\lambda_0 > |\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n| \,, \tag{4.11}$$

and a corresponding dominant eigenvector $\mathbf{b}_0$ with strictly positive components, $b_{i0} > 0 \,\forall\, i = 1, \ldots, n$.[5] In terms of components the differential equation in $\boldsymbol{\zeta}$ has the solutions

$$\zeta_k(t) \;=\; \zeta_k(0)\, \exp(\lambda_k\, t) \;. \tag{4.12}$$

Transformation back into the variables $\mathbf{z}$ yields

$$z_j(t) \;=\; \sum_{k=0}^{n-1} b_{jk}\, \beta_k(0)\, \exp(\lambda_k\, t) \;, \tag{4.13}$$

with the initial conditions encapsulated in the equation

$$\beta_k(0) \;=\; \sum_{i=1}^{n} h_{ki}\, z_i(0) \;=\; \sum_{i=1}^{n} h_{ki}\, x_i(0) \;. \tag{4.14}$$

From here we obtain the solutions in the original variables x_j through inverse transformation and normalization

$$x_j(t) \;=\; \frac{\sum_{k=0}^{n-1} b_{jk}\, \beta_k(0)\, \exp(\lambda_k\, t)}{\sum_{i=1}^{n} \sum_{k=0}^{n-1} b_{ik}\, \beta_k(0)\, \exp(\lambda_k\, t)} \;. \tag{4.15}$$

For sufficiently long times the contribution of the largest eigenvalue dominates the summations and we find for the long time solutions

$$\begin{aligned} x_j(t) \;&\approx\; \frac{b_{j0}\, \beta_0(0)\, \exp(\lambda_0\, t)}{\sum_{i=1}^{n} b_{i0}\, \beta_0(0)\, \exp(\lambda_0\, t)} \quad \text{and} \\ \lim_{t\rightarrow\infty} x_j(t) \;&=\; \bar{x}_j \;=\; \frac{b_{j0}\, \beta_0(0)}{\sum_{i=1}^{n} b_{i0}\, \beta_0(0)} \;, \end{aligned} \tag{4.16}$$

[5] We introduce here an asymmetry in numbering rows and columns in order to point at the special properties of the largest eigenvalue λ_0 and the dominant eigenvector $\mathbf{b}_0$.

which represent the components of the stationary population vector denoted by $\overline{\boldsymbol{\Upsilon}} = (\bar{x}_1, \ldots, \bar{x}_n)$. Solution curves $\boldsymbol{\Upsilon}(t) = \big(x_1(t), \ldots, x_n(t)\big)$ and stationary mutant distributions $\overline{\boldsymbol{\Upsilon}}$ were derived here in terms of the eigenvectors of the matrix W and can be obtained by numerical computation. Highly flexible and accurate computer codes for matrix diagonalization are available and the only problem that the numerical approach visualizes is high-dimensionality since n is commonly very large.

Perron-Frobenius theorem comes in two versions [259], which we shall now apply to the selection-mutation problem. The stronger version provides a proof for six properties of the largest eigenvector of non-negative primitive matrices[6] T:

(i) The largest eigenvalue is real and positive, $\lambda_0 > 0$,

(ii) a strictly positive right eigenvector $\boldsymbol{\ell}_0$ and a strictly positive left eigenvector $\mathbf{h}_0$ are associated with λ_0,

(iii) $\lambda_0 > |\lambda_k|$ holds for all eigenvalues $\lambda_k \neq \lambda_0$,

(iv) the eigenvectors associated with λ_0 are unique up to constant factors,

(v) if $0 \leq \mathrm{B} \leq \mathrm{T}$ is fulfilled and β is an eigenvalue of B, then $|\beta| \leq \lambda_0$, and, moreover, $|\beta| = \lambda_0$ implies $\mathrm{B} = \mathrm{T}$,

(vi) λ_0 is a simple root of the characteristic equation of T.

The weaker version of the theorem holds for irreducible matrices[7] T. All the above given assertions hold except (iii) has to be replaced by the weaker statement

(iii) $\lambda_0 \geq |\lambda_k|$ holds for all eigenvalues $\lambda_k \neq \lambda_0$.

Irreducible cyclic matrices can be used straightforwardly as examples in order to demonstrate the existence of conjugate complex eigenvalues. Perron-Frobenius theorem, in its strict or weaker form, holds not only for strictly

[6] A square non-negative matrix $\mathrm{T} = \{t_{ij}; i, j = 1, \ldots, n;\, t_{ij} \geq 0\}$ is called *primitive* if there exists a positive integer m such that T^m is strictly positive: $\mathrm{T}^m > 0$ which implies $\mathrm{T}^m = \{t_{ij}^{(m)}; i, j = 1, \ldots, n; t_{ij}^{(m)} > 0\}$.

[7] A square non-negative matrix $\mathrm{T} = \{t_{ij}; i, j = 1, \ldots, n; t_{ij} \geq 0\}$ is called *irreducible* if for every pair (i, j) of its index set there exists a positive integer $m_{ij} \equiv m(i, j)$ such that $t_{ij}^{m_{ij}} > 0$. An irreducible matrix is called *cyclic* with period d, if the period of (all) its indices satisfies $d > 1$, and it is said to be acyclic if $d = 1$.

positive matrices $\mathrm{T} > 0$ but also for large classes of mutation or value matrices ($\mathrm{W} \equiv \mathrm{T}$ being a primitive or an irreducible non-negative matrix) with off-diagonal zero entries corresponding to zero mutation rates. The occurrence of a non-zero element $t_{ij}^{(m)}$ in T^m implies the existence of a mutation path $\mathbf{X}_j \to \mathbf{X}_k \to \ldots \to \mathbf{X}_l \to \mathbf{X}_i$ with non-zero mutation frequencies for every individual step. This condition is almost always fulfilled in real systems.

The stationary solutions of the mutation-selection Equ. (4.9) represent the genetic reservoirs of an asexually reproducing populations or species, and have been called *quasispecies* therefore. The quasispecies is defined upon sequence space or upon a subspace of sequence space (see chapter 3). The condition for its existence in the sense of Perron-Frobenius theorem requires that every sequence can be reached from every other sequence along a finite-length path of single point mutations in the mutation network.[8] In addition the quasispecies contains all mutants at nonzero concentrations ($\bar{x}_i > 0 \,\forall\, i = 1, \ldots, n$). In other words, after sufficiently long time a kind of *mutation equilibrium* is reached at which all mutants are present in the population. In absence of neutral variants and in the limit of small mutation rates (see chapter 5) the quasispecies consists of a *master sequence*, the fittest sequence $\mathbf{X}_m : \{f_m = \max(f_i;\, i = 1, \ldots, n)\}$, and its mutants, $\mathbf{X}_j$ ($j = 1, \ldots, n, i \neq m$), which are present at concentrations that are, in essence, determined by their own fitness f_j, the fitness of the master f_m and the off-diagonal element of the mutation matrix Q_{jm} that depends on the Hamming distance[9] from the master sequence $d_{\mathrm{H}}(\mathbf{X}_j, \mathbf{X}_m)$ $\big($see Equ.(4.23)$\big)$. The coefficient of the first term in Equ. (4.9) for any sequence $\mathbf{X}_j$ can be partitioned into two parts: (i) the selective value $W_{jj} = Q_{jj} \cdot f_j$ and (ii) the mutational flow $\Omega_j = \sum_{i=1, i\neq j}^{n} Q_{ji} \cdot f_i$. For the master sequence $\mathbf{X}_j$ the two quantities, the

[8] By definition of fitness values, $f_i \geq 0$, and by definition of mutation frequencies, $Q_{ji} \geq 0$ where in both cases the greater sign ($>$) is valid for at least in one species, W is a non-negative matrix and the reachability condition boils down to the condition: $\mathrm{W}^k \gg 0$, i.e. there exists a k such that W^k has exclusively positive entries and Perron-Frobenius theorem applies [259].

[9] The Hamming distance named after Richard Hamming counts the number of positions in which two aligned sequences differ. The appropriate alignment of two sequences requires knowledge of their functions [215]. Here, we shall be concerned only with the simplest case: end-to-end alignment of sequences of equal lengths.

selective value of the master, W_{mm} and the *mutational backflow* Ω_m are of particular importance as we shall see in the discussion of *strong quasispecies*. At sufficiently large mutation rates it may also happen that another relatively fit sequence replaces the original master sequence, because its mutational backflow overcompensates the fitness difference. This replacement concerns not only the master sequence but also its mutation cloud: one quasispecies is replaced by another one (subsection 5.2.2).

Quasispecies in the sense of stationary mutant distribution are experimentally observed and can be quantitatively predicted in evolution of RNA in the test tube [14, 15]. Mutant distribution in viroid and virus populations, and in bacteria and can be subjected to quantitative analysis. The majority of investigations were done on RNA viruses [52] and these studies revealed that the question of stationarity of mutant distribution in the sense of a quasispecies cannot be confirmed. Because of the high mutation rates and the strong selection pressure exerted by the host's immune system the populations may never reach stationary states [53, 66, 149]. Modeling evolution by means of ODEs raises also several other questions that don't have trivial answers and are touching general problems concerning the conventional present view of evolution to which we shall come back in chapter 8.

4.3.2 Quasispecies and optimization

The selection problem has been illustrated in Fig.1.4 by means of trajectories on the unit simplex $\mathbb{S}_3^{(1)}$. We recall that – independently of the dimension of the system – all corners of the unit simplex,

$$\boldsymbol{e}_j = (x_j = 1; x_i = 0 \,\forall\, i = 1, \ldots, n; i \neq j) \ ,$$

were stationary states and all of them were saddle points except $\boldsymbol{e}_m$ corresponding to the sequence $\mathbf{X}_m$ with the highest fitness, $f_m = \max\{f_j; j = 1, \ldots, n\}$, which represents the only asymptotically stable point of the system.[10] Considering the position of $\boldsymbol{e}_m$ on the linear surface of mean fitness

[10] Precisely speaking, the corner $\boldsymbol{e}_l$ corresponding to the sequence $\mathbf{X}_l$ with the lowest fitness, $f_l = \min\{f_j; j = 1, \ldots, n\}$, is not a saddle point but a source, unless one takes into account as well the mode of convergence towards the simplex.

$\phi(\boldsymbol{x}) = \sum_{i=1}^{n} f_i x_i$ we recognize that maximal fitness is, of course, not the highest point of the fitness surface that extends to infinity but the highest point on the intersection of the physically acceptable subset $\mathbb{S}_n^{(1)}$ of $\mathbb{R}^{(n)}$ and fitness surface $\phi(\boldsymbol{x})$. Highest (or lowest) points of this kind are often called *corner equilibria*.

In order to consider the optimization problem in the selection-mutation case, we choose the eigenvectors of W as the basis of a new coordinate system shown in Fig. 4.6:

$$\boldsymbol{x}(t) \;=\; \sum_{i=1}^{n} x_k(t)\,\boldsymbol{e}_i \;=\; \sum_{k=0}^{n-1} \xi_k(t)\cdot \boldsymbol{b}_k \;,$$

wherein the vectors $\boldsymbol{e}_i$ are the unit eigenvectors of the conventional Cartesian coordinate system and $\boldsymbol{b}_k$ the eigenvectors of W. The unit eigenvectors represent the corners of $\mathbb{S}_n^{(1)}$ and in complete analogy we denote the space defined by the vectors $\boldsymbol{b}_k$ as $\tilde{\mathbb{S}}_n^{(1)}$. Formally, the transformed differential equation

$$\frac{\mathrm{d}\xi_k}{\mathrm{dt}} \;=\; \xi_k\,(\lambda_k - \phi)\,,\;\; k = 0, 1, \ldots, n-1 \;\;\text{with}\;\; \phi = \sum_{k=0}^{n-1} \lambda_k \xi_k = \overline{\lambda}$$

is identical to Equ. (1.14) and hence the solutions are the same,

$$\xi_k(t) \;=\; \xi_k(0)\,\exp\left(\lambda_k\, t \;-\; \int_0^t \phi(\tau)\,d\tau\right)\;,\;\; k = 0, 1, \ldots, n-1\;,$$

as well as the maximum principle **on** the simplex $\tilde{\mathbb{S}}_n^{(1)}$

$$\frac{d\phi}{dt} \;=\; \sum_{k=0}^{n-1} \frac{\mathrm{d}\xi_k}{\mathrm{dt}}\,\lambda_k \;=\; \sum_{k=0}^{n-1} \xi_k\,\lambda_k\;(\lambda_k - \phi) \;=\; <\lambda^2> \;-\; <\lambda>^2 \geq\; 0\;. \quad (1.20a)$$

The difference between the representation of selection and selection-mutation comes from the fact that the simplex $\tilde{\mathbb{S}}_n$ does not coincide with the physically defined space $\mathbb{S}_n$ (see Fig. 4.6 for a low-dimensional example). Indeed only the dominant eigenvector $\boldsymbol{b}_0$ lies in the interior of $\mathbb{S}_n^{(1)}$: It represent the stable stationary distribution of genotypes called quasispecies [70] towards which the solutions of the differential Equ. (4.9) converge. All other $n-1$ eigenvectors, $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{n-1}$ lie outside $\mathbb{S}_n^{(1)}$ in the *not physical* range where one

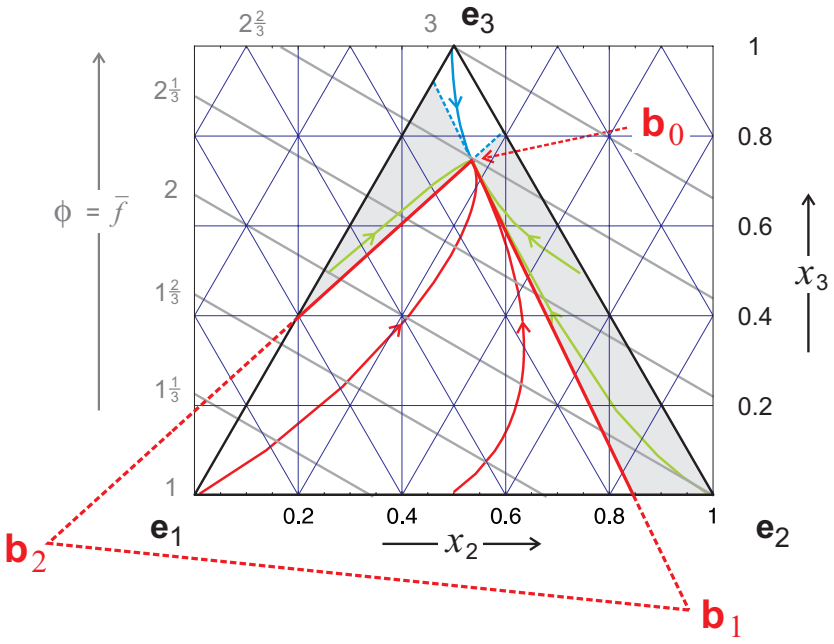


Figure 4.6: **The quasispecies on the unit simplex.** Shown is the case of three variables (x_1, x_2, x_3) on $\mathbb{S}_3^{(1)}$. The dominant eigenvector, the quasispecies denoted by $\boldsymbol{b}_0$, is shown together with the two other eigenvectors, $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$. The simplex is partitioned into an *optimization cone* (white; red trajectories) where the mean replication rate $\bar{f}(t)$ is optimized, a second zone, the *master cone* where $\bar{f}(t)$ always decreases (white; blue trajectory), and two other zones where may increase, decrease or change nonmonotonously (grey; green trajectories). In this illustration $\mathbf{X}_3$ is chosen to be the master sequence. Solution curves are presented as parametric plots $\boldsymbol{x}(t)$. In particular, the parameter values are: $f_1 = 1.9\,[\mathrm{t}^{-1}]$, $f_2 = 2.0\,[\mathrm{t}^{-1}]$, and $f_3 = 2.1\,[\mathrm{t}^{-1}]$, the Q-matrix was assumed to be bistochastic with the elements $Q_{ii} = 0.98$ and $Q_{ij} = 0.01$ for $i, j = \{1, 2, 3\}$. Then the eigenvalues and eigenvectors of W are:

k	λ_k	b_{1k}	b_{2k}	b_{3k}
1	2.065	0.093	0.165	0.742
2	1.958	0.170	1.078	-0.248
3	1.857	1.327	-0.224	-0.103

The mean replication rate $\bar{f}(t)$ is monotonously increasing along red trajectories, monotonously decreasing along the blue trajectory, and not necessarily monotonous along green trajectories.

or more variables x_i are negative. The quasispecies $\boldsymbol{b}_0$ is commonly dominated by a single genotype, the master sequence $\mathbf{X}_m$, having the largest stationary relative concentration: $\bar{x}_m \gg \bar{x}_i \,\forall\, i \neq m$, reflecting, for not too large mutation rates, the same sequence in the elements of the matrix W: $W_{mm} \gg W_{ii} \,\forall\, i \neq m$. As sketched in Fig. 4.6 the quasispecies is then situated close to the unit vector $\boldsymbol{e}_m$ in the interior of $\mathbb{S}_n^{(1)}$.

For the discussion of the optimization behavior the simplex is partitioned into three zones: (i) The zone of maximization of $\phi(t)$, the (large) lower white area in Fig. 4.6 where Equ. (1.20a) holds and which we shall denote as *optimization cone*,[11] (ii) the zone that includes the unit vector of the master sequence, $\boldsymbol{e}_m$, and the quasispecies, $\boldsymbol{b}_0$, as corners, and that we shall characterize as *master cone*,[11] and (iii) the remaining part of the simplex $\mathbb{S}_n^{(1)}$ (two zones colored grey in Fig. 4.6). It is straightforward to proof that increase of $\phi(t)$ and monotonous convergence towards the quasispecies is restricted to the optimization cone [256]. From the properties of the selection Equ. (1.14) we recall and conclude that the boundaries of the simplex $\tilde{\mathbb{S}}_n^{(1)}$ are invariant sets. This implies that no orbit of the differential Equ. (4.9) can cross these boundaries. The boundaries of $\mathbb{S}_n^{(1)}$, on the other hand, are not invariant but have the restriction that they can be crossed exclusively in one direction: from outside to inside.[12] Therefore, a solution curve starting in the optimization cone or in the master cone will stay inside the cone where it started and eventually converge towards the quasispecies, $\boldsymbol{b}_0$.

In zone (ii), the master cone, all variables ξ_k except ξ_0 are negative and ξ_0 is larger than one in order to fulfill the $L^{(1)}$-norm condition $\sum_{k=0}^{n-1} \xi_k = 1$. In order to analyze the behavior of $\phi(t)$ we split the variables into two groups, ξ_0 the frequency of the quasispecies and the rest [256], $\{\xi_k;\, k = 1, \ldots, n-1\}$

[11] The exact geometry of the optimization cone or the master cone is a polyhedron that can be approximated by a pyramid rather than a cone. Nevertheless we prefer the inexact notion *cone* because it is easier to memorize and to imagine in high-dimensional space.

[12] This is shown easily by analyzing the differential equation, but follows also from the physical background: No acceptable process can lead to negative particle numbers or concentrations. It can, however, start at zero concentrations and this means the orbit begins at the boundary and goes into the interior of the physical concentration space, here the simplex $\mathbb{S}_n^{(1)}$.

with $\sum_{k=1}^{n-1}\xi_k = 1-\xi_0$:

$$\frac{d\phi}{dt} \;=\; \lambda_0^2\xi_0 \,+\, \sum_{k=1}^{n-1}\lambda_k^2\xi_k \,-\, \left(\lambda_0\xi_0 + \sum_{k=1}^{n-1}\lambda_k\xi_k\right)^2 \;.$$

Next we replace the distribution of λ_k values in the second group by a single λ-value, $\tilde{\lambda}$ and find:

$$\frac{d\phi}{dt} \;=\; \lambda_0^2\xi_0 \,+\, \tilde{\lambda}^2(1-\xi_0) \,-\, \Bigl(\lambda_0\xi_0 + \tilde{\lambda}(1-\xi_0)\Bigr)^2 \;.$$

After a view simple algebraic operations we find eventually

$$\frac{d\phi}{dt} \;=\; \xi_0\,(1-\xi_0)\,(\lambda_0-\tilde{\lambda})^2 \;. \tag{4.17}$$

For the master cone with $\xi_0 \geq 1$, this implies $d\phi(t)/dt \leq 0$, the flux is a non-increasing function of time. Since we are only interested in the sign of $d\phi/dt$, the result is exact, because we could use the mean value $\tilde{\lambda} = \bar{\lambda} = (\sum_{k=1}^{n-1}\lambda_k\xi_k)/(1-\xi_0)$, the largest possible value λ_1 or the smallest possible value λ_{n-1} without changing the conclusion. Clearly, the distribution of λ_k-values matters for quantitative results. As it has to be, Equ. (4.17) applies also to the optimization cone and gives the correct result that $\phi(t)$ is non-decreasing. Decrease of mean fitness or flux $\phi(t)$ in the master cone is readily illustrated: Consider, for example, a homogeneous population of the master sequence as initial condition: $x_m(0) = 1$ and $\phi(0) = f_m$. The population becomes inhomogeneous because mutants are formed. Since all mutants have lower replication constants by definition, $(f_i < f_m\,\forall\, i \neq m)$, ϕ becomes smaller. Finally, the distribution approaches the quasispecies $\boldsymbol{b}_0$ and $\lim_{t\to\infty}\phi(t) = \lambda_0 < f_m$.

An extension of the analysis from the master cone to the grey zones, where not all ξ_k values with $k \neq 0$ are negative, is not possible. It has been shown by means of numerical examples that $d\phi(t)/dt$ may show nonmonotonous behavior and can go through a maximum or a minimum at finite time [256].

4.3.3 Error thresholds

How many mutations per generation can be tolerated without jeopardizing inheritance? This is a proper question that is very hard to analyze without

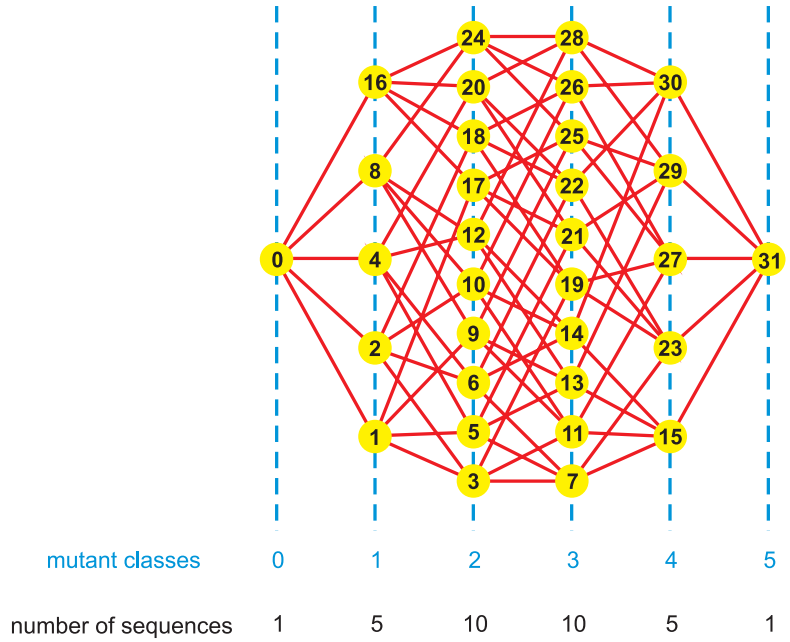


Figure 4.7: **Mutant classes in sequence space.** The sketch shows the sequence space for binary sequences of chain length $l = 5$, which are given in terms of their decadic encodings: "**0**" ≡ **00000**, "**1**" ≡ **00001**, ..., "**31**" ≡ **11111**. All pairs of sequences with Hamming distance $d_{\mathrm{H}} = 1$ are connected by red lines. The number of sequences in mutant class k is $\binom{l}{k}$.

a theory that has a direct access on the dependence of mutant distribution on the mutation frequency. For conventional population genetics it is not simple to give an answer, because mutations are not part of the evolutionary dynamics modeled. In the theory based on chemical kinetics of replication, however, mutation is just another replication channel and propagation of errors is part of the system. Here, an analytical expression for the stationary mutant distribution – the quasispecies $\bar{\Upsilon}$ – as a function of the error rate will be provided by means of the zero mutational backflow approximation. A limit for error propagation, which is compatible with evolution, is derived and the results are compared with perturbation theory and accurate numerical results. Eventually we present a proof for the existence of a phase transition-like error threshold.

Neglect of mutational backflow approximation. Neglect of mutational backflow from mutants to the master sequence allows for the derivation of analytical approximations for the quasispecies [65, 67]. The backflow is of the form

$$\Phi_{m\leftarrow(i)} \;=\; \sum_{i=1}^{n} Q_{mi} f_i \bar{x}_i \;=\; \sum_{i=1}^{n} W_{mi}\, \bar{x}_i \;,$$

and if $W_{mi} << |W_{mm} - W_{ii}|\,(i \neq m)$ is fulfilled $Q_{mi} = 0\,\forall\, i = 1, \ldots, n; i \neq m$ is a valid approximation for small mutation rates [280]. Insertion into equation (4.9) yields the following ODE for the master sequence[13]

$$\frac{\mathrm{d}x_m^{(0)}}{\mathrm{dt}} \;=\; (W_{mm} \,-\, \phi)\, x_m^{(0)} \;=\; (Q_{mm}\, f_m \,-\, \phi)\, x_m^{(0)} \;. \tag{4.18a}$$

$$\frac{\mathrm{d}x_j^{(0)}}{\mathrm{dt}} \;=\; W_{jm}\, x_m^{(0)} - \phi\, x_j^{(0)} \;=\; Q_{jm}\, f_m\, x_m^{(0)} \,-\, \phi\, x_j^{(0)} \;. \tag{4.18b}$$

The differential equation (4.18a) sustains two stationary states: (i) $\bar{x}_m^{(0)} = 0$, the state of extinction, and (ii) $Q_{mm} f_m - \phi = 0$. In the latter case we split ϕ as we did previously:

$$\phi \;=\; \sum_{i=1}^{n} f_i\, x_i^{(0)} \;=\; f_m\, x_m^{(0)} \,+\, \sum_{i=1, i\neq m}^{n} f_i\, x_i^{(0)} \;=\; f_m\, x_m^{(0)} \,+\, (1 - x_m^{(0)})\, \overline{f}_{-m}$$

$$\text{with} \quad \overline{f}_{-m} \;= \frac{1}{1 - x_m} \sum_{i=1, i\neq m}^{n} f_i\, x_i \;.$$

Insertion into condition (ii) yields

$$Q_{mm}\, f_m \,-\, f_m\, \bar{x}_m^{(0)} \,-\, (1 - \bar{x}_m^{(0)})\, \overline{f}_{-m} \;,$$

which can be evaluated to yield an expression for $\bar{x}_m^{(0)}$. For known concentrations of the master sequence we obtain the concentration of the mutants from equation (4.18b). For simplicity we introduce the assumption of the single peak landscape leading to $\overline{f}_{-m} = f$:

$$\bar{x}_j^{(0)} \;=\; \frac{Q_{jm}\, f_m\, \bar{x}_m^{(0)}}{(f_m - f)\, \bar{x}_m^{(0)} \,+\, f} \;.$$

[13] The superscript '(0)' stands for zeroth order perturbation theory and means (total) neglect of mutational backflow, although the approach does not correspond to a defined order of perturbation theory (see next paragraph).

which after some algebraic operations leads to an equation for the stationary concentrations of all members of the quasispecies

$$\bar{x}_m^{(0)} = \frac{W_{mm} - \overline{f}_{-m}}{f_m - \overline{f}_{-m}} = \frac{Q_{mm} - \sigma_m^{-1}}{1 - \sigma_m^{-1}} , \tag{4.19a}$$

$$\bar{x}_j^{(0)} = \varepsilon^{d_{i0}}\, \bar{x}_m^{(0)} ;\ j = 1, \ldots, n\,,\ j \neq m \tag{4.19b}$$

$$\text{with}\ \ \sigma_m = f_m \,/\, \overline{f}_{-m} = f_m \,/\, f \ \ \text{and}\ \ \varepsilon = \frac{p}{1-p}.$$

The superiority σ_m is a measure of the advantage in fitness the master has over the rest of the population, and $\overline{f}_{-m}$ is the mean fitness of this rest.[14] In case of the single peak fitness landscape we have the trivial result: $\overline{f}_{-m} = f$. The superiority of the master sequence can also be understood as an empirical quantity that can be determined through direct measurements of replication efficiencies of cloned sequences.

In order to gain basic insight into evolutionary dynamics we introduce the uniform error rate model. Any polynucleotide sequence is represented by a string, $\mathbf{X}_j = \big(s_1^{(j)}, s_2^{(j)}, \ldots, s_n^{(j)}\big)$ with $s_1^{(j)} \in \{\mathbf{A}, \mathbf{U}, \mathbf{G}, \mathbf{C}\}$. The uniform error model assumes that the *point mutation rate* $p\,(s_i^{(j)})$ is independent of the site s_i, the nucleobase where the mutation goes ($\mathbf{A}\rightarrow\mathbf{U}$, $\mathbf{A}\rightarrow\mathbf{G}$ or $\mathbf{A}\rightarrow\mathbf{C}$, for example), and the particular sequence $\mathbf{X}_j$. We can interpret the uniform error rate model, for example, by the application of a mean mutation rate per site and reproduction event p. As a matter of fact this assumption is often violated: There exist, for example, *hot spots for mutation* where mutations are substantially more frequent than in other regions of the sequence. In general, however, the uniform error rate model turned out to be extremely useful because it is sufficiently realistic and at the same time accessible to rigorous mathematical analysis. Further simplification is introduced by the use of binary rather than four-letter sequences.[15] Then, diagonal and off-

[14] An exact calculation of $\overline{f}_{-m}$ is difficult because it requires knowledge of the stationary concentrations of all variants in the population: $\bar{x}_i$; $x = 1, \ldots, n$. For computational details see [67, 68, 71, 250].

[15] It should be noted that artificially synthesized two letter (DU; D = 2,6-diaminopurine) ribozymes have perfect catalytic properties [237].

diagonal elements of matrix Q are of the simple form

$$\begin{aligned} Q_{ii} &= (1-p)^l \quad \text{and} \\ Q_{ji} &= (1-p)^{l-d_{\mathrm{H}}(\mathbf{X}_j,\mathbf{X}_i)}\, p^{\,d_{\mathrm{H}}(\mathbf{X}_j,\mathbf{X}_i)} = (1-p)^l \varepsilon^{d_{\mathrm{H}}(\mathbf{X}_j,\mathbf{X}_i)} \\ &\text{with } \varepsilon = \frac{p}{1-p}\ , \end{aligned} \tag{4.20}$$

wherein the Hamming distance $d_{\mathrm{H}}(\mathbf{X}_i, \mathbf{X}_j)$ counts the numbers of positions in which the two aligned sequences $\mathbf{X}_l$ and $\mathbf{X}_j$ differ. It is worth noticing that the Hamming distance is symmetric $d_{\mathrm{H}}(\mathbf{X}_i, \mathbf{X}_j) = d_{\mathrm{H}}(\mathbf{X}_j, \mathbf{X}_i)$ and this is one of the three properties required for a metric:

(i) $d_{\mathrm{H}}(\mathbf{X}_i, \mathbf{X}_j) = 0$,

(ii) $d_{\mathrm{H}}(\mathbf{X}_i, \mathbf{X}_j) = d_{\mathrm{H}}(\mathbf{X}_j, \mathbf{X}_i)$ (symmetry), and

(iii) $d_{\mathrm{H}}(\mathbf{X}_i, \mathbf{X}_j) + d_{\mathrm{H}}(\mathbf{X}_j, \mathbf{X}_k) \geq d_{\mathrm{H}}(\mathbf{X}_i, \mathbf{X}_k)$ (triangle inequality).

Indeed, the Hamming distance induces a metric in sequence space (chapter 3). Equ. (4.20) allows for a calculation of all elements of the mutation matrix Q from three parameters only: (i) the point mutation rate p, (ii) the lengths of the sequences expressed in the number of nucleotides l, and (iii) the Hamming distance d_{H}. For given sequences of the same length l and d_{H} are known and p is the only free parameter apart from the fitness values f_j.

The uniform error rate allows for further simplification of the mutation-selection equation (4.9) since the corresponding mutation matrix Q is consistent with a change in variables from individual sequences to mutant classes (Fig. 4.7) [224, 280]. The sequences are ordered in mutant classes $\mathbf{Y}_j$ according to the Hamming distance from the master sequence, which is commonly the sequence with the highest fitness value. In absence of neutrality the zero-error class contains only the master sequence $\big(\mathbf{Y}_0 : \{\mathbf{X}_m \equiv \mathbf{X}_0\}\big)$, the one-error class comprises all single point mutations, the two-error class all double point mutations, and so on.[16] Since the error rate p is independent

[16] Centering the quasispecies around the master sequence suggests to use decadic equivalents, e.g., $6 \equiv \mathbf{110}$, for the individual (binary) sequences: $\mathbf{X}_0 = \mathbf{000}\ldots\mathbf{00}$, $\mathbf{X}_1 = \mathbf{000}\ldots\mathbf{01}$, $\mathbf{X}_2 = \mathbf{000}\ldots\mathbf{10}, \cdots, \mathbf{X}_{2^l-1} = \mathbf{111}\ldots\mathbf{11}$.

of the sequence and because of the assumption of a single-peak fitness landscape all molecules belonging to the same mutant class have identical fitness values $f_{(k)} = \{f_0, f\}$, it is possible to introduce variables for entire mutant classes Γ_k (Fig. 4.7):

$$y_k \;=\; \sum_{j\,,\mathbf{X}_j\in\Gamma_k} x_j\;,\;\; k=0,1,\ldots,l\;,\;\sum_{k=0}^{l} y_k = 1\;. \tag{4.21}$$

The mutation matrix Q has to be adjusted to transitions between classes [224, 280]. For mutations from class Γ_l into Γ_k we calculate:

$$Q_{kj} \;=\; \sum_{i=j+k-l}^{\min(k,j)} \binom{k}{i}\binom{l-k}{j-i} p^{\,k+j-2i}(1-p)^{\,l-(k+j-2i)}\;. \tag{4.22}$$

The mutation matrix Q for error classes is not symmetric, $Q_{kj} \neq Q_{jk}$ as follows from Equ. (4.22).

Insertion into Equ. (4.19) yields the three parameter (l, p, σ) expression

$$\bar{x}_m \;\approx\; \frac{(1-p)^l - \sigma_m^{-1}}{1-\sigma_m^{-1}} \quad \text{and} \tag{4.23a}$$

$$\bar{x}_j \;\approx\; \varepsilon^{d_{\mathrm{H}}(\mathbf{X}_j,\mathbf{X}_m)}\frac{f_m}{f_m - f_j}\,\bar{x}_m\;. \tag{4.23b}$$

Equ. (4.23) provides a quantitative estimate for the concentrations of mutants: For given p, $\bar{x}_j$ is the larger the smaller the Hamming distance from the master, $d_{\mathrm{H}}(\mathbf{X}_j, \mathbf{X}_m)$, is and the smaller the larger the difference in fitness, $f_m - f_j$ is. The stationary concentration of the master sequence, $\bar{x}_m$, vanishes at some critical mutation rate $p = p_{\mathrm{cr}}$ characterized as *error threshold* or more dramatically as *error catastrophe* [16, 65, 68, 70, 280]:[17]

$$p_{\mathrm{cr}} \;=\; 1 \,-\, \sigma^{-1/l} \;\;\Longrightarrow\;\; p_{\max} \;\approx\; \frac{\ln\,\sigma}{l} \quad \text{and} \quad l_{\max} \;\approx\; \frac{\ln\,\sigma}{p}\;. \tag{4.24}$$

[17] Zero or negative concentrations of sequences clearly contradict the exact results described above and are an artifact of the zero backflow approach. As said before Perron-Frobenius theorem guarantees positive concentrations for all mutants that can be derived from the master sequence through a finite numbers of single point mutations. Nevertheless, the agreement between the exact solutions and the zero backflow results up to the error threshold as shown in Fig. 4.8 is remarkable.

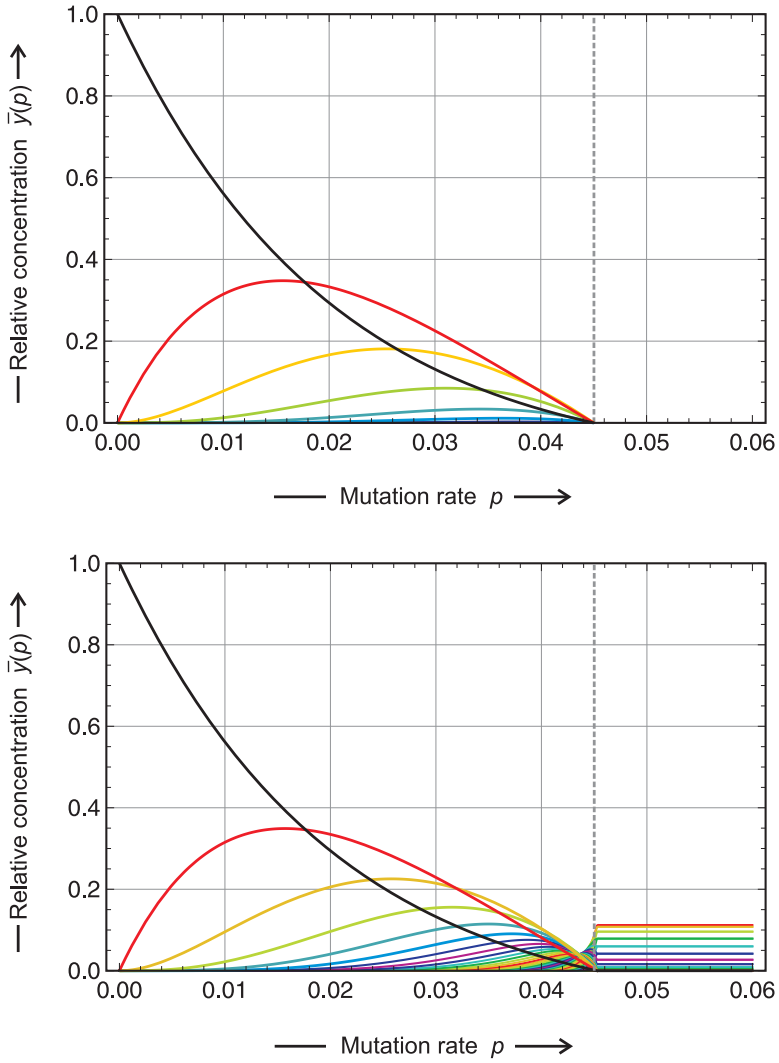


Figure 4.8: **The quasispecies as a function of the point mutation rate p.** The plot shows the stationary mutant distribution of sequences of chain length $l = 50$ on a single-peak fitness landscape as a function of the point mutation rate p. The upper part contains the approximation by the zero backflow approximation according to Equ. (4.23) and is compared with the exact results presented in the lower part of the figure. Plotted are the relative concentration of entire mutant classes (Fig. 4.7): $\bar{y}_0$ (black) is the master sequence $\mathbf{X}_m \equiv \mathbf{X}_0$, $\bar{y}_1$ (red) is the sum of the concentrations of all one-error mutants of the master sequence, $\bar{y}_2$ (yellow) that of all two-error mutants, $\bar{y}_3$ (green) that of all three-error mutants, and so on. In the zero backflow approach the entire population vanishes at a critical mutation rate p_{cr} called the error threshold (which is indicated by a broken gray line at $p_{\mathrm{cr}} = 0.04501$) whereas a sharp transition to the uniform distribution, Π, is observed with the exact solutions. In the uniform distribution the concentration of class k is given by $\binom{l}{k} \,/\, 2^l$ with a largest value of $\bar{y}_{25} = 0.1123$ and a smallest value of $\bar{y}_0 = \bar{y}_{50} = 8.8818 \times 10^{-16}$. Choice of parameters: $f_0 = f_m = 10$, $f = f_j = 1 \,\forall\, j = 1, \ldots, l;\ j \neq m$, and accordingly $\overline{f}_{-m} = 1$.

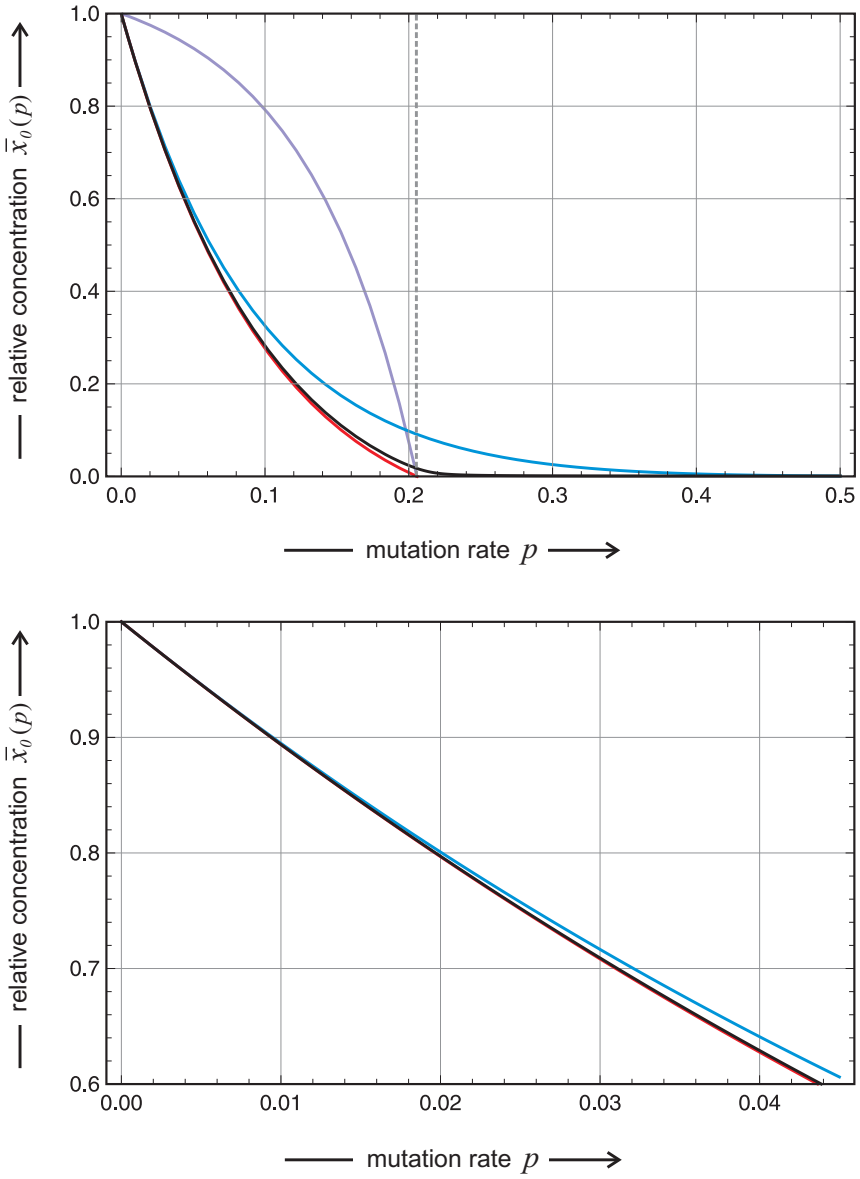


Figure 4.9: **Quasispecies calculated by perturbation theory.** The upper plot presents a comparison of the stationary concentration of the master sequence in the zero mutational backflow approximation, $\bar{x}_m^{(0)}(p)$ (red), with first order perturbation theory, $\bar{x}_m^{(1)}(p)$ (blue), and numerical solution $\bar{x}_m$ (black). The violet curve is the total concentration in the zero mutational backflow approximation, $\bar{c}^{(0)} = \sum_{i=1}^{n} \bar{x}_i^{(0)}$. The lower plot is an enlargement of the curves at low mutation rates demonstrating that both approximations almost coincide with the exact solution curve. The error threshold indicated by a broken vertical line occurs at $p = p_{\mathrm{cr}} = 0.2057$. Choice of parameters: $n = 10$, $f_0 = 10\,[\mathrm{t}^{-1}]$, and $f = 1\,[\mathrm{t}^{-1}]$.

Fig. 4.8 compares a quasispecies calculated by the zero backflow approach with the exact solution and shows excellent agreement up to the critical mutation rate $p = p_{\rm cr}$. This agreement is important because quantitative applications of quasispecies theory to virology are essentially based on Eqs. (4.23) and (4.24) [54].

Zero mutational backflow fails to account for the quasispecies at mutation rates above the threshold $p_{\rm cr}$: Perron-Frobenius theorem states that the concentrations of all members of the quasispecies are positive definite: $\bar{x}_i > 0 \,\forall\, i = 1, \ldots, n$ but zero mutational backflow yields $\bar{x}_i = 0 \,\forall\, i = 1, \ldots, n$ at $p = p_{\rm cr}$. Considering the problem more closely this is no surprise since the zero mutational backflow assumption violates the conditions for the validity of the theorem: The requirement for matrix W was irreducibility and this implies that every sequence can be reached from every other sequence in a finite number of mutation steps – zero mutational backflow implies that the master cannot be reached from the mutants. Beyond the error threshold we have to consider either full first order perturbation theory or the numerical solutions. The manipulation of the elements of the matrix Q has also the consequence that the stationary total concentration $\bar{c}^{(0)} = \sum_{i=1}^{n} \bar{x}_i^{(0)}$ is not constant but vanishes at the error threshold

$$\bar{c}^{(0)}(p) \;=\; \frac{1}{Q}\,\frac{Q - \sigma_m^{-1}}{1 - \sigma_m^{-1}} \;.$$

Clearly, the good agreement between the zero mutational backflow approximation and the exact solution is not fortuitous and many examples have shown that it is quite general and becomes perfect for long chains l (see *Proof for the existence of an error threshold* below).

Perturbation theory. Application of first and second order Rayleigh-Schrödinger perturbation theory to calculate the quasispecies $\bar{\Upsilon}$ as a function of the mutation rate has been performed in the past [280]. Here we present the full analytical first order expressions $\bar{x}_i^{(1)}(p)$. The second expressions are rather clumsy and bring only limited improvement for small mutation rates p.

The largest eigenvalue is the same by zero mutational backflow and in first order perturbation theory:

$$\lambda_0^{(0)} \;=\; \lambda_0^{(1)} \;=\; W_{mm} \;=\; Q_{mm}\, f_m \;.$$

For the computation of the largest eigenvector we make use of the first order expression from perturbation theory of the matrix W (4.19b):[18]

$$\bar{x}_j^{(1)} \;=\; \frac{W_{jm}}{W_{mm}-W_{jj}}\;\bar{x}_m^{(1)}\,;\; j=1,\ldots,n\,,\; j\neq m\;.$$

Making use of the normalization condition $\sum_{i=1}^{n}\bar{x}_i^{(1)}=1$ we obtain for the master sequence

$$\bar{x}_m^{(1)} \;=\; \frac{1}{1+\sum_{i=1,i\neq m}^{n}\frac{W_{im}}{W_{mm}-W_{ii}}}\;.$$

Straightforward calculations yield for the stationary concentrations

$$\bar{x}_m^{(1)}(p) \;=\; \frac{Q\,(1-\sigma_m^{-1})}{1-Q\,\sigma_m^{-1}}\;, \tag{4.25a}$$

$$\bar{x}_j^{(1)}(p) \;=\; \frac{Q_{im}}{Q\,(1-\sigma_m^{-1})}\,\bar{x}_m^{(0)}\,;\; j=1,\ldots,n\,,\; j\neq m \tag{4.25b}$$

$$\text{with}\;\; Q \;=\; (1-p)^{\nu}\;.$$

As shown in figure 4.9 the curve $\bar{x}_m^{(1)}(p)$ extends to the point $p=\tilde{p}=\frac{1}{2}$ and further, but it does not pass precisely through the uniform distribution Π,

$$\bar{x}_m^{(1)}(\frac{1}{2}) \;=\; \left(\frac{1}{2}\right)^{\nu}\;\frac{1-\sigma_m^{-1}}{1-(\frac{1}{2})^{\nu}\sigma_m^{-1}}\;.$$

The deviation of $\bar{x}_m^{(1)}$ from numerical solution is much larger than in the zero mutational backflow approximation approximation and the error threshold phenomenon is not detectable. In summary, first order perturbation theory provides a consistent approximation to the eigenvalues and eigenvectors of the value matrix W. The results, however, are not nearly as good as those of the zero back mutation approach. Improvements by second order are possible at very small error rates but the calculations are rather tedious and the solutions for the eigenvalue $\lambda_0^{(2)}$ become unstable for larger error rates [280]. A combination of zero mutation flux approximation and first order

[18] As a matter of fact, the first order perturbation expressions are used in the zero mutational backflow approximation for the calculation of the concentrations of the mutants, because

perturbation theory in the sense of equations (4.18a) and (4.25b) [65, 280] leads to slightly better results than the zero mutation flux approach alone but is not recommended because of the lack of consistency.

Numerical solutions. Full solutions can be computed numerically through solving the eigenvalue problem of matrix W for different values of of the mutation rate p, and for a typical example the normalized concentrations of error classes $\bar{y}_{(k)}(p)$ are shown in figure 4.8. The agreement between the numerical results and the zero mutational backflow curve for the master class, $\bar{y}_0(p)$ in the region above the error threshold is remarkable indeed. The other solution curves $\bar{y}_{(k)}(p)$ $(k \neq 0)$ agree well too but the deviations become larger with increasing k.

The numerical solution for the master sequence (black curve) decreases monotonously from $p = 0$ to $p = \tilde{p} = 1/2$, this is between two points for which analytical solutions exist. At vanishing error rates, $\lim p \to 0$, the master sequence is selected, $\lim_{t\to\infty} x_0(t) = \lim_{t\to\infty} y_0(t) = \bar{y}_0 = \bar{x}_0 = 1$, and all other error classes vanish in the long time limit. Increasing error rates are reflected by a decrease in the stationary relative concentration of the master sequence and a corresponding increase in the concentration of all mutant classes. Except $\bar{y}_0(p)$ all concentrations $\bar{y}_k(p)$ with $k < \nu/2$ go through a maximum at values of p that increases with k – as in case of zero mutational backflow where we had an implicit analytical expression for the maximum, and approach the curves for $\bar{y}_{\nu-k}$ – whereas the zero mutational backflow curves still go through a maximum because they vanish at $p = p_{\mathrm{cr}}$. At $p = \tilde{p} = 1/2$ we have $\tilde{p} = 1 - \tilde{p}$ for binary sequences, and again the eigenvalue problem can be solved exactly (see section 4.3.6. The uniform distribution Π is a result of the fact that correct digit incorporation and point mutation are equally probable for binary sequences at $\tilde{p} = 1/2 = 1 - \tilde{p}$ and therefore we may characterize this scenario as random replication.[19] It is worth mentioning that the range of high mutation rates $\tilde{p} \leq p \leq 1$ is also meaningful: At $p = 1$ the complementary digit is incorporated with

[19] The extension to sequences over an alphabet with κ classes of digits is straightforward. In the frame of uniform errors random replication occurs at $\tilde{p} = 1/\kappa = (1 - \bar{p})/(\kappa - 1)$. For the natural four letter alphabet we have $\tilde{p} = 1/4$.

ultimate accuracy, $\mathbf{0} \to \mathbf{1}$ and $\mathbf{1} \to \mathbf{0}$, and accordingly, the range at high p-values describes error-prone complementary or plus-minus replication [280].

The special situation at the error threshold is the occurrence of an (almost) uniform distribution far away from the point $p = \tilde{p}$ – in figure 4.8 the critical mutation rate is $p_{\text{cr}} = 0.045 \ll \tilde{p} = 0.5$. As we shall see in the next section 4.3.6, the error threshold on the single peak landscape is characterized by the coincidence of three phenomena: (i) the concentration of the master sequence becomes very small and this is expressed in term of level crossing values $\bar{y}_0(p)|_{p=p_{(1/M)}} = 1/M$ where M is 100, 1000 or higher depending on the size of 2^ν, (ii) a sharp change in the quasispecies distribution within a narrow band of p-values that reminds of a phase transition [187, 188, 267, 282, 298], and (iii) a transition to the uniform distribution, which implies that the domain within which the uniform distribution is fulfilled to a high degree of accuracy has the form of a broad plateau ($p_{\text{cr}} = 0.045 < p < \tilde{p} = 0.5$ in figure 4.8). It is worth considering the numerical data from the computations shown in the figure: $p_{\text{cr}} = 0.04501$ from zero mutational backflow versus the level crossing values $p_{(1/100)} = 0.04360$, $p_{(1/1000)} = 0.04509$, and $p_{(1/10000)} = 0.04525$.

Proof for the existence of an error threshold. In order to present a rigorous proof for the existence of an error threshold on the single peak landscape in the sense that the exact solution converges to the zero mutational back flow result in the limit of infinite chain length l. Previously we stated that the agreement between the (exact) numerical solution for the stationary quasispecies and the zero mutational backflow in surprisingly good and here we shall give a rigorous basis for this agreement. The proof proceeds in three steps: (i) Models are derived that provide upper and lower bounds for the exact solution, (ii) the models are evaluated analytically in order to yield expressions for the relative stationary concentration of the master sequence $\mathbf{X}_m$ at the position of the error threshold, $\bar{x}_m^{(\text{flow})}(p_{\text{cr}})$, and (iii) we show that the values for the upper and the lower bound coincide in the limit $l \to \infty$.

The zero mutational backflow approximation neglects backflow completely; now we introduce two other approximations that are based on model back-

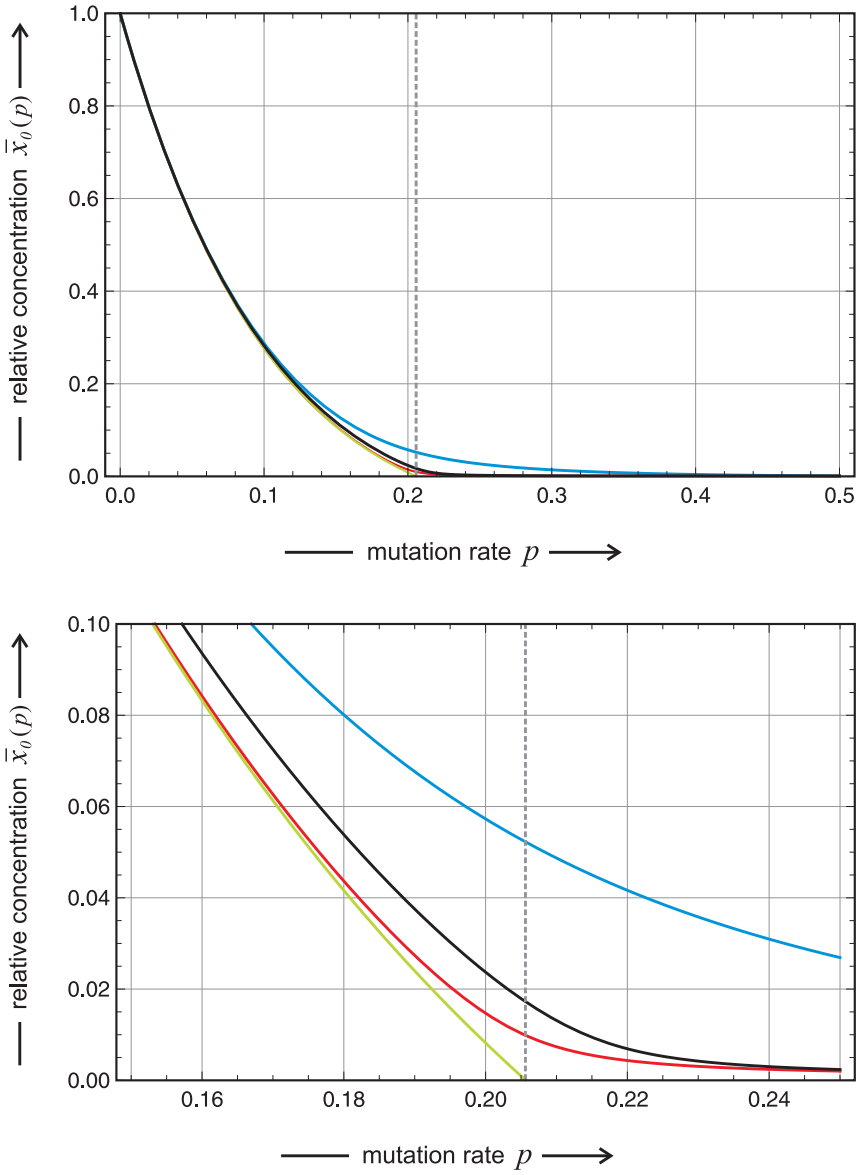


Figure 4.10: **Existence of the error threshold.** The plots represent the exact solution (black) together with the zero mutational backflow approximation (green), the uniform backflow approximation (red) and the error-class one backflow approximation (red). The numerically exact solution is entrapped between the uniform and the one error-class approximation. Since both approximations converge to zero in the limit of long chain lengths ($l \to \infty$) the exact curve does as well. The error threshold as indicated by a broken vertical line occurs at $p = p_{\mathrm{cr}} = 0.2057$. Choice of parameters: $n = 10$, $f_0 = 10\,[\mathrm{t}^{-1}]$, and $f = 1\,[\mathrm{t}^{-1}]$.

flows that represent lower and upper bounds for the exact backflow. Computation of the mutational backflow requires either knowledge of the distribution of concentrations of all sequence of an assumption about it. In order to be able to handle the problem analytically the backflow must lead to an autonomous equation for the master concentration x_0. The minimal backflow can be estimated by the assumption of a uniform distribution (Π) for all sequences except the master. In this case all sequences contribute equally no matter whether a particular sequence is close to the master sequences or far apart. For the concentrations $x_i = (1 - x_0^{(\Pi)})/(n-1)\,\forall\, i = 1, \ldots, n$ with $n = \kappa^l$ we obtain under the further assumption of a single peak landscape and the uniform error rate model the ODE for $x_0^{(\Pi)}$:

$$\begin{aligned} \frac{\mathrm{d}x_0^{(\Pi)}}{\mathrm{dt}} &= x_0^{(\Pi)}(Q_{00}f_0 - \phi) + f\,\frac{1 - x_0^{(\Pi)}}{n-1} \\ \text{with } \phi &= f + (f_0 - f)\,x_0^{(\Pi)} \end{aligned} \tag{4.26}$$

The stationary concentration is obtained as the solution of a quadratic equation

$$\bar{x}_0^{(\Pi)} = \frac{Qf_0 - f - f\gamma(1-Q) + \sqrt{\Big(Qf_0 - f - f\gamma(1-Q)\Big)^2 + 4(f_0 - f)(1-Q)f\gamma}}{2(f_0 - f)}$$

$$\text{with } \mathrm{Q} = \mathrm{Q}_{00} = (1-\mathrm{p})^{\mathrm{l}} \text{ and } \gamma = \frac{1}{\mathrm{n}-1}$$

Insertion of the value of the mutation rate parameter at the error threshold, $p = p_{\mathrm{cr}} = 1 - \sigma^{-1/l}$ or $Q = (1 - p_{\mathrm{cr}})^l = \sigma^{-1}$ leads to the result

$$\bar{x}_0^{(\Pi)}(p_{\mathrm{cr}}) = \frac{1}{2}\,\frac{\sqrt{1 + 4\sigma(n-1)} - 1}{\sigma(n-1)}\,, \tag{4.27}$$

which yields in the limit of long chains or large l-values

$$\bar{x}_0^{(\Pi)}(p_{\mathrm{cr}}) \approx \frac{1}{\sqrt{\sigma\, n}} = \frac{1}{\sqrt{\sigma\,\kappa^l}}\,. \tag{4.27'}$$

Ultimately the value of the stationary concentration of the master sequence decays exponentially with one halt of the chain lengths as exponent: $\bar{x}_0^{(\Pi)}(p_{\mathrm{cr}}) \propto \kappa^{-l/2}$. It is straightforward to show that the uniform mutational backflow approximation becomes exact at the point $p = \tilde{p}$ and insertion in the quadratic equation yields:

$\bar{x}_0^{(\Pi)}(\frac{1}{2}) = \left(\frac{1}{2}\right)^l$. Generalization, of course, is straightforward and yields $p = \tilde{p}$ and insertion in the quadratic equation yields: $\bar{x}_0^{(\Pi)}(1/\kappa) = \left(1/\kappa\right)^l$

In order to find an upper bound for the stationary solution of the master sequence we assume that mutational backflow comes only form the sequences in the one error class, $\Gamma_1^{(0)}$, which can be assumed to be present at equal concentrations, $x_i = (1 - x_0^{(\mathbf{I})})/l$, and $\sum_{i=1}^{l} x_i = 1 - x_0$. All other sequences except the master sequence and the one error class are absent. Pointing at the fact that $\Gamma_1^{(0)}$ represents the entire mutant cloud we shall denote this distribution by **I**. For the corresponding elements of the mutation matrix Q we use the usual expressions, which are all equal: $Q_{0i} = Q_{0(1)} = Q_{01} \,\forall\, i = 1, \ldots, l$. The ODE for the master sequence is then again autonomous and can be readily solved for the stationary state:

$$\begin{aligned} \frac{\mathrm{d}x_0^{(\mathbf{I})}}{\mathrm{dt}} &= x_0^{(\mathbf{I})}(Q_{00}f_0 - \phi) + Q_{01}\, f\, (1 - x_0^{(\mathbf{I})}) \\ &\text{with} \quad \phi = f + (f_0 - f)\, x_0^{(\mathbf{I})} \; . \end{aligned} \tag{4.28}$$

The stationary concentration is again obtained from a quadratic equation of similar structure as before

$$\bar{x}_0^{(\mathbf{I})} = \frac{Q_{00}f_0 - Q_{01}f - f + \sqrt{\Big(Q_{00}f_0 - Q_{01}f - f\Big)^2 + 4(f_0 - f)Q_{01}f}}{2(f_0 - f)}$$

$$\text{with} \quad Q_{00} = (1-p)^l \quad \text{and} \quad Q_{01} = (1-p)^{l-1}\, p \; .$$

It is shown straightforwardly that the curve for the class one backflow passes through the point $p = \tilde{p} = \kappa^{-l}$. For the stationary concentration of the master sequence at the error threshold we find

$$\bar{x}_0^{(\mathbf{I})}(p_{\mathrm{cr}}) = -\frac{Q_{01}\, f}{2\,(f_0 - f)} + \sqrt{\frac{Q_{01}\, f}{f_0 - f}} \cdot \sqrt{1 + \frac{Q_{01}\, f}{4\,(f_0 - f)}} \; , \tag{4.29}$$

with three components. Before we can discuss the individual terms we have to examine the asymptotic dependence of the mutation rate p on the chain length l, which is encapsulated in the series expansion

$$Q_{01} = (1-p)^{l-1}\, p = p - (l-1)\, p^2 + \frac{(l-1)(l-2)}{2}\, p^3 - \cdots$$

with the first term being p. The critical mutation rate can be approximated by $p_{\mathrm{cr}} \approx \ln \sigma / l$ and we can consider Equ. (4.29). The negative term in equation shows

Table 4.1: **Proof of the error threshold.** Two auxiliary model assumptions concerning mutational backflow are applied: (i) uniform distribution for all one error mutants, called *class one uniform*, and (ii) the uniform distribution Π for all sequences except the master sequence denoted as *all uniform*. All solution curves $\bar{x}_0(p)$ – exact and approximate – begin at the value $\bar{x}_0(0) = 1$ and all except the zero backflow approximation end at the point $\bar{x}(\tilde{p}) = \bar{x}(\kappa^{-1}) = 1/\kappa^l$

		$\bar{x}_0$-value at the mutation rate		
mutational backflow	notation	$p = 0$	$p = p_{\text{cr}}$	$\tilde{p} = \frac{1}{\kappa}$
class one uniform	$\bar{x}_0^{(\mathbf{I})}(p)$	1	$\sqrt{\ln\sigma/(\sigma-1)}\cdot 1/\sqrt{l}$	κ^{-l}
exact	$\bar{x}_0(p)$	1	computed	κ^{-l}
all uniform	$\bar{x}_0^{(\Pi)}(p)$	1	$1/\sqrt{\sigma\kappa^l}$	κ^{-l}
zero	$\bar{x}_0^{(0)}(p)$	1	0	negative

an asymptotic dependence on the chain length of l^{-1}, the first factor behaves asymptotically like $1/\sqrt{l}$ whereas the second factor converges to unity. What remains in the limit of long chains or large l-values is

$$\bar{x}_0^{(\mathbf{I})}(p_{\text{cr}}) \;\approx\; \sqrt{\frac{f\,\ln\sigma}{f_0-f}}\cdot\frac{1}{\sqrt{l}} \;=\; \sqrt{\frac{\ln\sigma}{\sigma-1}}\cdot\frac{1}{\sqrt{l}}\;. \tag{4.29'}$$

The value of the stationary concentration of the master sequence decays with the reciprocal square root of the chain length: $\bar{x}_0^{(\mathbf{I})}(p_{\text{cr}}) \propto 1/\sqrt{l}$. Although the class one uniform distribution is not an impressively good upper bound for the exact solution curve, it is sufficient for our purpose here because $\bar{x}_0^{(\mathbf{I})}(p_{\text{cr}})$ vanishes in the limit $l \to \infty$.

In summary the solution curve of the mutation-selection equation (4.9) for the master sequence and the three approximations at the critical mutation rate p_{cr} appear in the order shown in Tab. 4.1 and there is no reason to doubt that the same order prevails for the entire domain $0 < p < \tilde{p} = \kappa^{-l}$: The exact solution is indeed entrapped between the two approximations for the mutational backflow and, since both converge asymptotically to zero the exact curve approaches the zero mutational backflow approximation in the limit of long chains. All four curves (Fig. 4.10) start at the point $\bar{x}_m(0)$ and all except the zero backflow approximation end at the correct value $\tilde{p} = \kappa^{-l}$.

Table 4.2: **Mutation rates and genome length.** The table shows the product between the mutation rate per site and copying event, p, and the genome length l, which is roughly constant within of class of organisms [58].

class of organisms	mutation rate per genome $\mu = p \cdot l$	reproduction event
RNA viruses	1	replication
retroviruses	0.1	replication
bacteria	0.003	cell division
eukaryotes	0.003	cell division
eukaryotes	0.01 – 0.1	sexual reproduction

The stationary master concentrations at the critical mutation rate $p = p_{\text{cr}}$ illustrate the relative importance of the mutational backflow (Fig. 4.29): Zero backflow assumption causes the stationary concentration $\bar{x}_0^{(0)}$ to vanish. The all uniform backflow is a little more than one half of the computed exact value, and this approximation is a excellent lower bound for the exact solution. The error one class backflow is about three time as large as the exact solution. Nevertheless it is an upper bound for the real mutation flow and serves the purpose for which it is intended here. If one is interested in an approximation apart from this proof the zero mutational back flow approximation $\bar{x}_0^{(\Pi)}(p)$ in the region $0 \leq p < p_{\text{cr}}$ and the uniform backflow approximation in the entire range are suitable approximations. In particular the uniform backflow approximation is well suited because it is exact for $p = 0$ and $p = \tilde{p} = 1/\kappa$ and it has also a correct asymptotic behavior in the long chain limit at the error threshold.

Error thresholds and applications. The error threshold has been put in relation to phase transitions [187, 188, 282]. Here, we are in a position to prove the behavior of the exact solution curve $\bar{x}_0(p)$ in the limit $l \to \infty$. The critical mutation rate converges to the value zero: $\lim_{l\to\infty} p_{\rm cr} = \lim_{l\to\infty} \ln\sigma/l = 0$. At the same time we have $\lim_{l\to\infty} \bar{x}_0 = 0$ for $p > 0$ and thus the quasispecies degenerates to an "L"-shaped distribution, $\bar{x}_0(0) = 1$ and $\bar{x}_0(p) = 0\,\forall\, p > 0$, and we are left with a pathological phase transition at $p_{\rm cr} = 0$.

According to Equ. (4.24) the error threshold defines a maximal error rate for evolution, $p \leq p_{\max}$, at constant chain length l, and at constant reproduction accuracy p the length of faithfully copied polynucleotides is confined to $l \leq l_{\max}$ [70, 72]. The first limit of a maximal error rate $p_{\max}$ has been used in pharmacology for the development of new antiviral strategies [56], and the second limit entered hypothetical modeling of early biological evolution where the accuracy limits of enzyme-free replication confine the lengths of polynucleotides that can be replicated faithfully [73].

The error threshold relation (4.24) can be written in a different form that allows for straightforward testing with empirical data:

$$\mu \;=\; l\,p \;\approx\; \ln\sigma \;, \tag{4.24'}$$

the product of the genome length and the mutation per site and replication, μ, which represents the mean number of mutations per full genome reproduction, corresponds to the logarithm of the superiority. In a publication by John Drake *et al.* the mutation rates μ were found to be remarkably constant for organisms of the same class (table 4.2 and [58]). In other words, for organisms within one class – viruses, retroviruses, bacteria, and eukaryotes – the replication is more accurate if the genome is longer. A comparison between the bacteriophage Q$\beta$and the vesicular stomatitis virus (SVS) may serve as an example [57]: The genome lengths are 4 200 and 11 200 and the mutation rates per site and replication are 1.5×10^{-3} and 3.8×10^{-4}, respectively. The large difference in reproduction accuracy between mitosis and meiosis is remarkable.

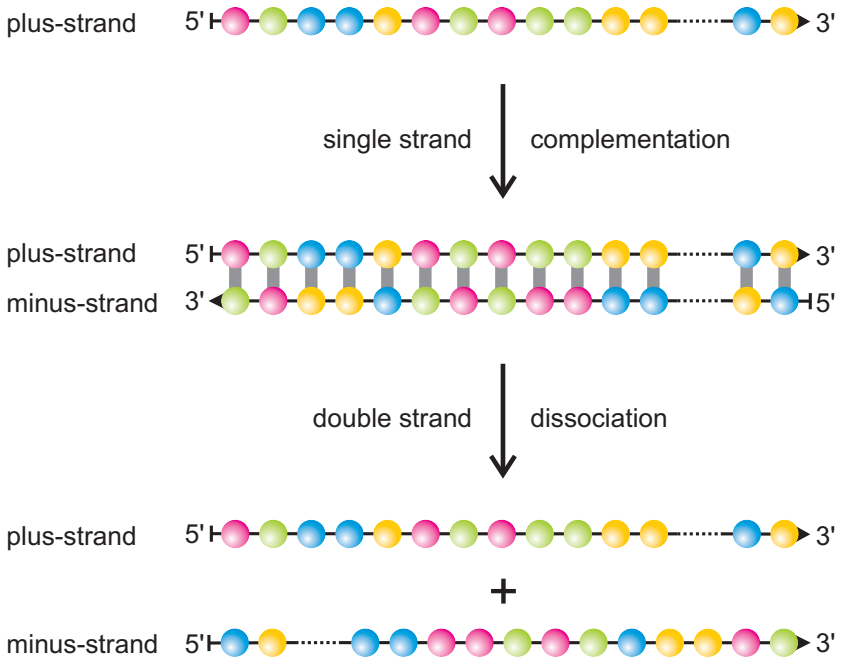


Figure 4.11: **The principle of complementary replication**. Complementary replication involves two recurrent logical steps: (i) the synthesis of a double strand or duplex from a single strand, say the plus-strand, and (ii) duplex dissociation into the newly synthesized minus-strand and the template plus-strand. In the next replication round the minus-strand is the template for plus-strand synthesis. Plus- and minus-strand together grow exponentially just like a simple replicator would.

4.3.4 Complementary replication

The molecular mechanism of replication of RNA single strands in the test tube or in virus infected cells proceeds through an intermediate represented by an RNA molecule with the complementary sequence (Fig. 4.11). Here we denote the plus-strand by $\mathbf{X}_1 \equiv \mathbf{X}^{(+)}$ and the complementary minus-strand by $\mathbf{X}_2 \equiv \mathbf{X}^{(-)}$ and the corresponding rate parameters by f_1 and f_2, respectively:

$$\begin{aligned} (\mathbf{A}) + \mathbf{X}_1 &\xrightarrow{f_1} \mathbf{X}_2 + \mathbf{X}_1 \quad \text{and} \\ (\mathbf{A}) + \mathbf{X}_2 &\xrightarrow{f_2} \mathbf{X}_1 + \mathbf{X}_2 \,. \end{aligned} \tag{4.30}$$

In analogy to Equ. (4.8), with $f_1 = k_1\,[\mathbf{A}]$, $f_2 = k_2\,[\mathbf{A}]$, $x_1 = [\mathbf{X}_1]$, $x_2 = [\mathbf{X}_2]$, and $x_1 + x_2 = 1$ we obtain the following differential equation [65, 70, 71]:

$$\begin{aligned} \frac{\mathrm{d}x_1}{\mathrm{dt}} &= f_2 x_2 - x_1\,\phi \quad \text{and} \\ \frac{\mathrm{d}x_2}{\mathrm{dt}} &= f_1 x_1 - x_2\,\phi \quad \text{with} \quad \phi = f_1 x_1 + f_2 x_2\ . \end{aligned} \tag{4.31}$$

Applying again the integrating factor transformation [329, p. 322ff.] yields the linear equation

$$\frac{\mathrm{d}z_1}{\mathrm{dt}} = f_2 z_2 \ \text{ and } \ \frac{\mathrm{d}z_2}{\mathrm{dt}} = f_1 z_1 \ \text{ or } \ \frac{\mathrm{d}\boldsymbol{z}}{\mathrm{dt}} = \mathrm{W}\cdot\boldsymbol{z};\ \boldsymbol{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix},\ \mathrm{W} = \begin{pmatrix} 0 & f_2 \\ f_1 & 0 \end{pmatrix}.$$

The eigenvalues and (right hand) eigenvectors of the matrix W are

$$\begin{aligned} \lambda_{1,2} &= \pm\sqrt{f_1 f_2} = \pm f \quad \text{with} \quad f = \sqrt{f_1 f_2}\ , \\ \boldsymbol{b}_1 &= \begin{pmatrix} \sqrt{f_2} \\ \sqrt{f_1} \end{pmatrix} \quad \text{and} \quad \boldsymbol{b}_2 = \begin{pmatrix} -\sqrt{f_2} \\ \sqrt{f_1} \end{pmatrix}\ . \end{aligned} \tag{4.32}$$

Straightforward calculation yields analytical expressions for the two variables (see paragraph mutation) with the initial concentrations $x_1(0)$ and $x_2(0)$, and $\gamma_1(0) = \sqrt{f_1}x_1(0) + \sqrt{f_2}x_2(0)$ and $\gamma_2(0) = \sqrt{f_1}x_1(0) - \sqrt{f_2}x_2(0)$ as abbreviations:

$$\begin{aligned} x_1(t) &= \frac{\sqrt{f_2}\left(\gamma_1(0)\cdot e^{ft} + \gamma_2(0)\cdot e^{-ft}\right)}{(\sqrt{f_1}+\sqrt{f_2})\gamma_1(0)\cdot e^{ft} - (\sqrt{f_1}-\sqrt{f_2})\gamma_2(0)\cdot e^{-ft}} \\ x_2(t) &= \frac{\sqrt{f_1}\left(\gamma_1(0)\cdot e^{ft} - \gamma_2(0)\cdot e^{-ft}\right)}{(\sqrt{f_1}+\sqrt{f_2})\gamma_1(0)\cdot e^{ft} - (\sqrt{f_1}-\sqrt{f_2})\gamma_2(0)\cdot e^{-ft}}\ . \end{aligned} \tag{4.33}$$

After sufficiently long time the negative exponential has vanished and we obtain the simple result

$$x_1(t) \to \sqrt{f_2}/(\sqrt{f_1}+\sqrt{f_2})\,,\ x_2(t) \to \sqrt{f_1}/(\sqrt{f_1}+\sqrt{f_2}) \ \text{ as } \ \exp(-ft) \to 0\,.$$

After an initial period, the plus-minus pair, the ensemble $\mathbf{X}_{1,2} \equiv \mathbf{X}^{(\pm)}$, grows like a single replicator with a fitness value $f = \sqrt{f_1 f_2}$ and a stationary ratio of the concentrations of complementary stands $\bar{x}_1/\bar{x}_2 \approx \sqrt{f_2}/\sqrt{f_1}$. It is worth noticing that the faster replicating strand is present at a lower equilibrium concentration: $\bar{x}_1\sqrt{f_2} = \bar{x}_2\sqrt{f_1}$. By mass action it is guaranteed that the ensemble grows at an optimal rate.

4.3.5 Quasispecies on "simple" model landscapes

First we consider the general properties of a quasispecies distribution of binary sequences as a function of the mutation rate $\overline{\Upsilon}(p)$. There are two p-values for which we have exact analytical solutions: (i) error-free replication at $p = 0$, and (ii) random replication at $p = \frac{1}{2}$. The first case is the selection scenario that we have discussed already in section 1.2: All sequences except the master sequence vanish after sufficiently long time, $\overline{\Upsilon}(0) = (\bar{x}_m = 1, \bar{x}_j = 0 \,\forall\, j \neq m)$.

For $p = \frac{1}{2}$ the incorporation of the incorrect digits into the growing binary string has exactly the same probability as the incorporation of the correct digit, $1 - p = \frac{1}{2}$, wrong and right digits are chosen at random and every sequence has the same probability to be outcome of a replication event – as easily visualized, the template plays no role. Accordingly, the mutation matrix Q and the value matrix W are of the form:

$$\mathrm{Q} = \left(\frac{1}{2}\right)^l \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}, \quad \mathrm{W} = \left(\frac{1}{2}\right)^l \begin{pmatrix} f_1 & f_2 & \dots & f_n \\ f_1 & f_2 & \dots & f_n \\ \vdots & \vdots & \ddots & \vdots \\ f_1 & f_2 & \dots & f_n \end{pmatrix}.$$

Matrix W has only one nonzero eigenvalue $\lambda_0 = (\frac{1}{2})^l \sum_{i=1}^n f_i$ the number of sequences $n = 2^l$, and the corresponding eigenvector is the uniform distribution: $\overline{\Upsilon}(\frac{1}{2}) = 2^{-l}(1, 1, \dots, 1)$. Between these two limiting cases the quasispecies fulfills $\sum_{i=1}^n \bar{x}_i = 1$ and $\bar{x}_i > 0 \,\forall\, i$ by Perron-Frobenius theorem. The influence of the landscape on the solution curves $\overline{\Upsilon}(p)$ is thus restricted to the way the transition from the homogeneous ($p = 0$) to the uniform distribution occurs. We shall call the transition smooth when there is no recognizable abrupt change in at some critical value of the mutation rate parameter $p = p_{\mathrm{cr}}$ (for an example see Fig. 4.12). A typical sharp transition in $\overline{\Upsilon}(p)$ from a structured distribution to the uniform distribution is shown in Fig. 4.8: It occurs in a very narrow range of p around p_{cr}, and $p_{\mathrm{cr}} = 0.045$ is far away from the random replication point ($p = \frac{1}{2}$). Thus we have a wide range of p-values where $\overline{\Upsilon}(p)$ (almost) coincides with the uniform distribu-

tion.

At first we consider the quasispecies $\overline{\Upsilon}(p)$ for the entire range of mutation rates $0 \leq p \leq 1$. As we have discussed above, the left half of the domain, $0 \leq p \leq \frac{1}{2}$ describes (direct) error-prone replication and ends at the uniform distribution when the probability to make an error at the incorporation of a single digit (p) is exactly as large as the probability to incorporate the correct digit ($1-p$). What happens if the mutation rate parameter exceeds one half? At $p = 1$ always the opposite digit is incorporated: $\mathbf{0} \to \mathbf{1}$ and $\mathbf{1} \to \mathbf{0}$, and we are dealing with complementary replication.[20] Accordingly, we expect pairs of complementary sequences to be target of selection in the right half of the domain, $\frac{1}{2} \leq p \leq 1$. At $p = 1$ a master pair is selected and as we saw in subsection 4.3.4 the ratio of the two sequences is $\bar{x}_+/\bar{x}_- = \sqrt{f_-/f_+}$, and all other sequences vanish. Between the three points where we have exact results solutions are readily obtained by numerical computation.

Next we are now in the position to study the conditions for the occurrence of an error threshold on the single peak landscape. In Figs. 4.12-4.15 we show results for the single peak landscape. In addition to the dependence of the stationary class concentrations, $\bar{y}_k(p)$, we show also the curves for the individual sequences $\bar{x}_j(p) = \bar{y}_k(p)/\binom{l}{k} \, \forall \, \mathbf{X}_j \in \Gamma_k$. A chain length of $l = 5$ is very small compared to $l = 50$ in Fig. 4.8 and we see no indication of an error threshold. The situation reminds of cooperative transitions [189, 324, 325] or phase transitions [187, 188, 282] in the sense that the transition becomes sharper with increasing sequence length l. There is, however, a second possibility to induce an error threshold in quasispecies dynamics through reducing the difference in fitness between the master sequence and the mutant cloud, $f_0 - f_n$. In the three plots these difference decreases from 9 (Fig. 4.12) to 3 (Fig. 4.13) and eventually to 0.1 (Fig. 4.14) and at the same time the transition changes from smooth to broad and then to sharp. In the curves $\bar{x}_j(p)$

[20] In real polynucleotide replication, the situation is a little bit more involved since the minus-strand is not the complement of the plus strand but the 3'end-5'end swapped complement (Fig. 4.11). If we consider the string $\mathbf{X}_0 = (\mathbf{000\cdots0})$ as plus strand, $\mathbf{X}_{2^l-1} = (\mathbf{111\cdots1})$ is the minus strand since swapping does not change palindromic sequences.

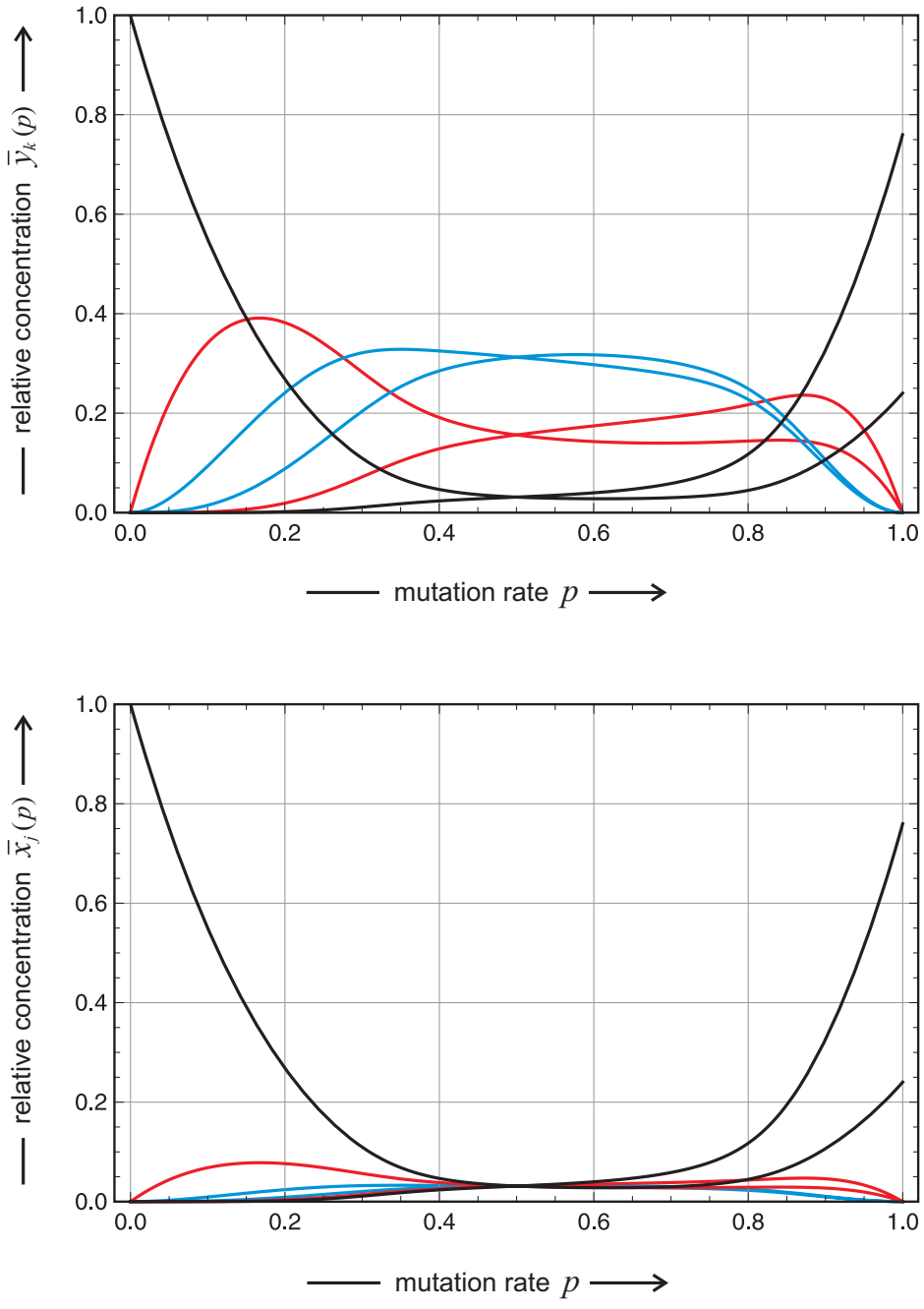


Figure 4.12: **The quasispecies on the single peak landscape I**. The plots show the dependence of the quasispecies from binary sequences on the mutation rate parameter $\overline{\Upsilon}(p)$ over the full range $0 \leq p \leq 1$. The upper plot shows the relative concentrations of entire mutant classes, $\bar{y}_k$ $(k = 0, 1, \ldots, l)$, the lower plot presents the relative concentrations of individual sequences, $\bar{x}_j$ $(j = 0, \ldots, 2^l - 1)$. Choice of parameters: $l = 5$, $f_0 = 10$, and $f_n = 1$.

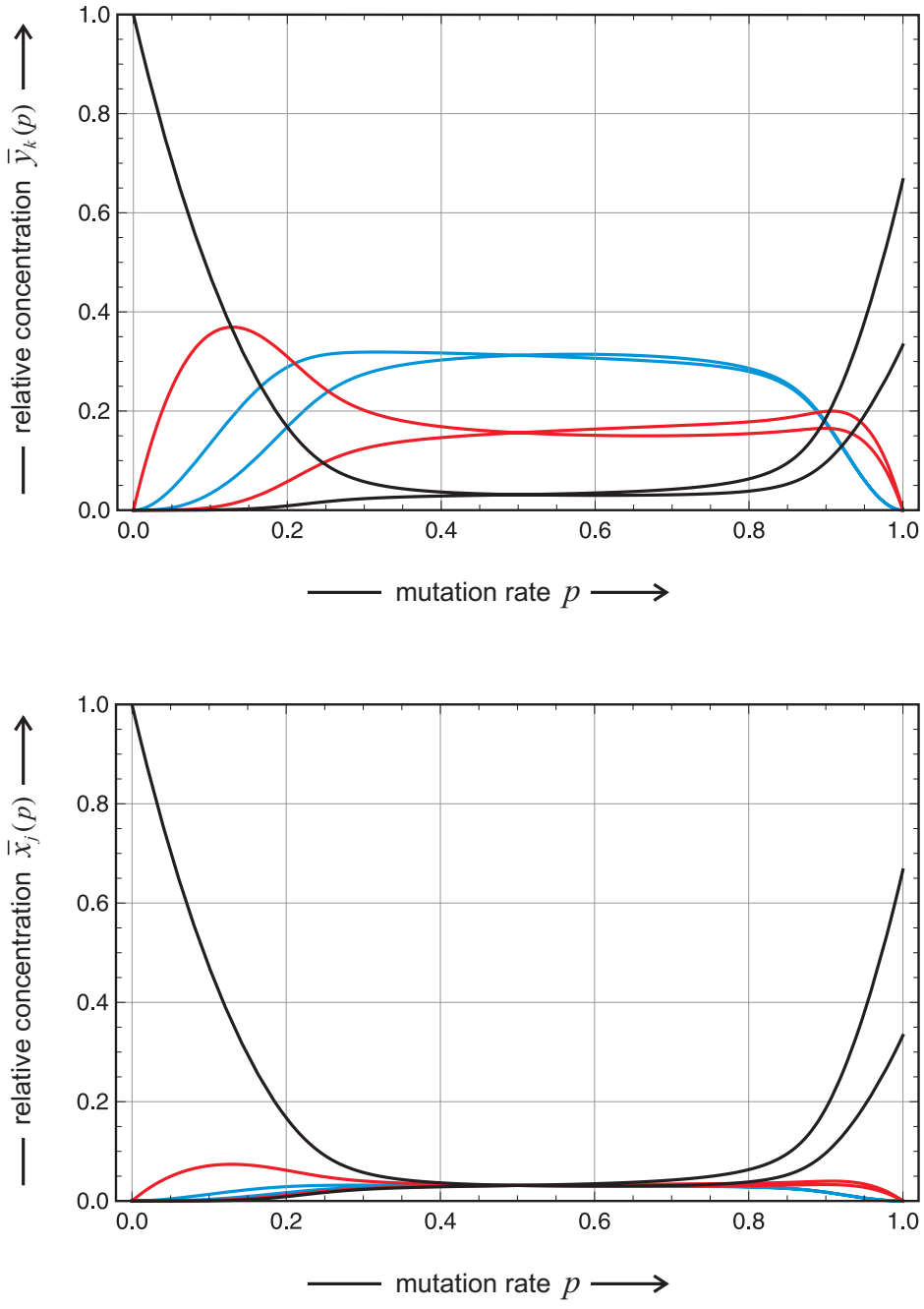


Figure 4.13: **The quasispecies on the single peak landscape II**. The plots show the dependence of the quasispecies from binary sequences on the mutation rate parameter $\overline{\Upsilon}(p)$ over the full range $0 \leq p \leq 1$. The upper plot shows the relative concentrations of entire mutant classes, $\bar{y}_k$ $(k = 0, 1, \ldots, l)$, the lower plot presents the relative concentrations of individual sequences, $\bar{x}_j$ $(j = 0, \ldots, 2^l - 1)$. Choice of parameters: $l = 5$, $f_0 = 4$, and $f_n = 1$.

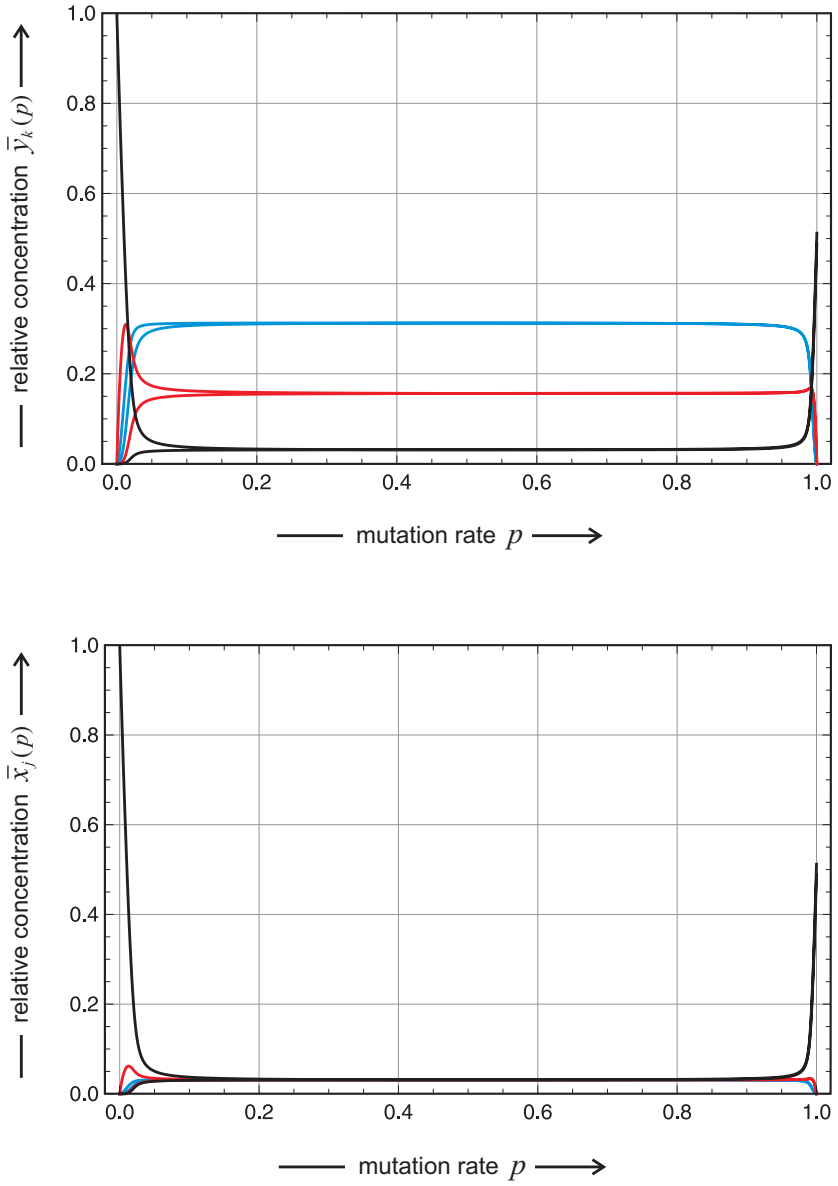


Figure 4.14: **The quasispecies on the single peak landscape III**. The plots show the dependence of the quasispecies from binary sequences on the mutation rate parameter $\overline{\Upsilon}(p)$ over the full range $0 \leq p \leq 1$. The upper plot shows the relative concentrations of entire mutant classes, $\bar{y}_k$ $(k = 0, 1, \ldots, l)$, the lower plot presents the relative concentrations of individual sequences, $\bar{x}_j$ $(j = 0, \ldots, 2^l - 1)$. Choice of parameters: $l = 5$, $f_0 = 1.1$, and $f_n = 1$.

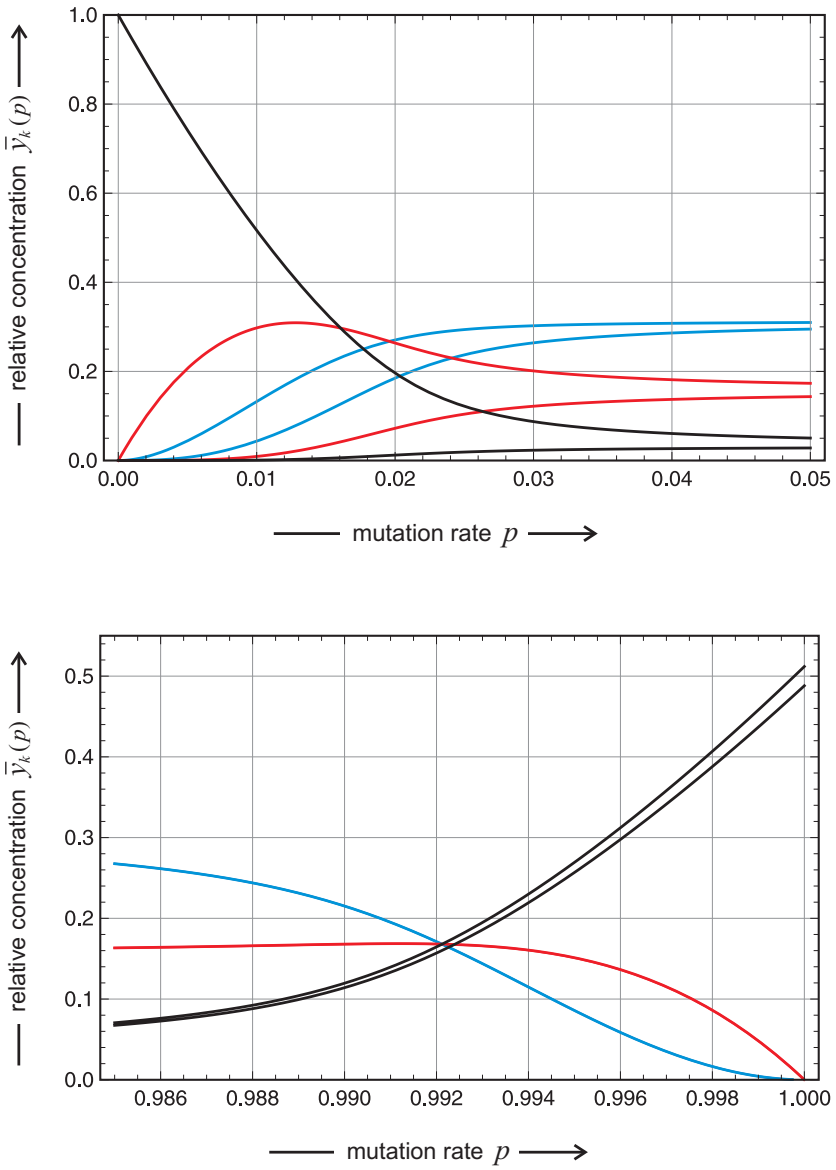


Figure 4.15: **The quasispecies on the single peak landscape IV**. The plots show the dependence of the quasispecies from binary sequences on the mutation rate parameter $\overline{\Upsilon}(p)$ over the full range $0 \leq p \leq 1$. The upper plot is an enlargement of the lhs of Fig. 4.14 and shows the relative concentrations of entire mutant classes, $\bar{y}_k$ $(k = 0, 1, \ldots, l)$. The lower plot enlarges the curves on the rhs. Choice of parameters: $l = 5$, $f_0 = 1.1$, and $f_n = 1$.

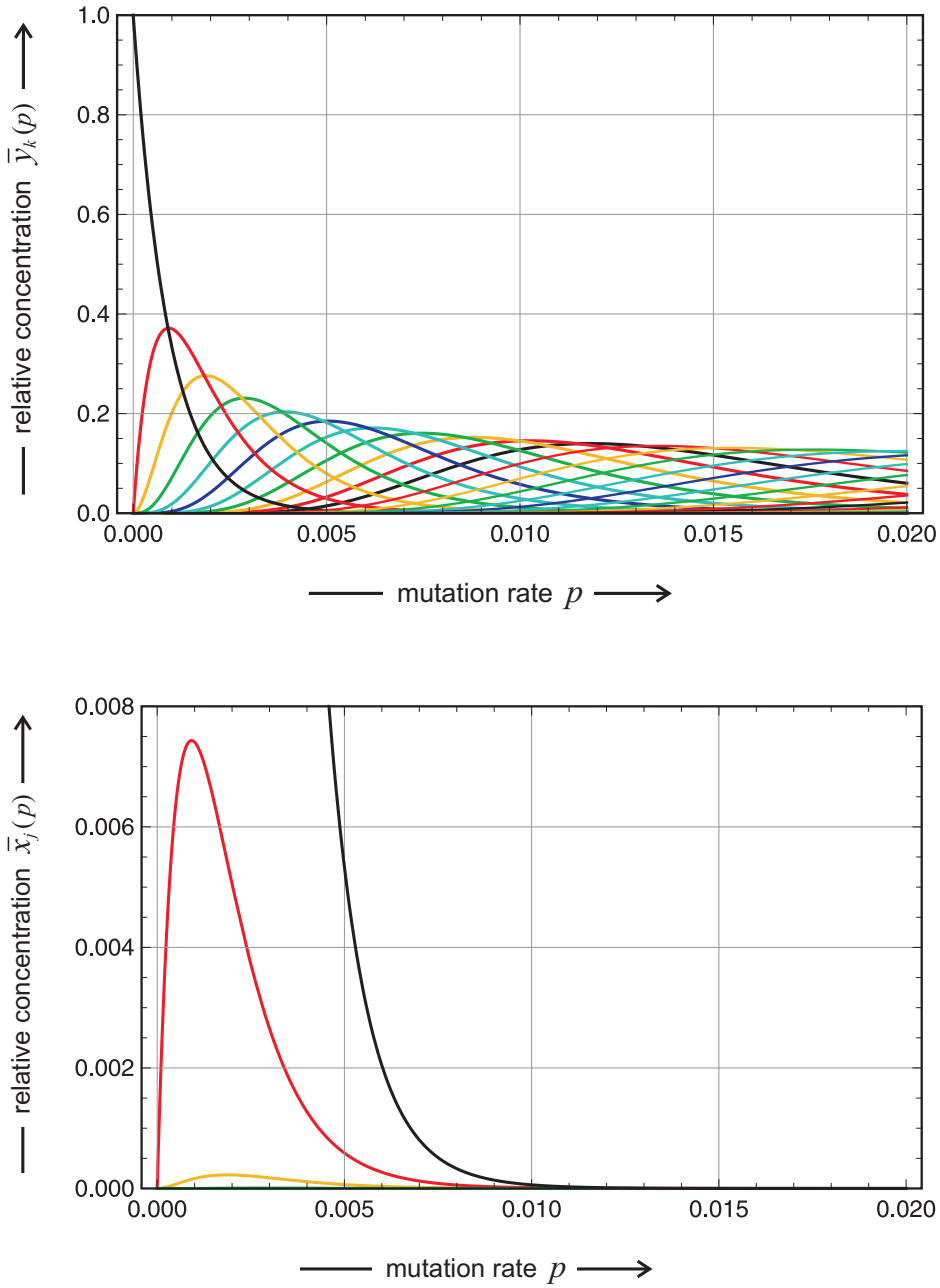


Figure 4.16: **The quasispecies on the multiplicative landscape**. The plots show the dependence of the quasispecies from binary sequences on the mutation rate parameter $\overline{\Upsilon}(p)$ on the multiplicative landscape (3.4b). The upper plot shows the relative concentrations of entire mutant classes, $\bar{y}_k(p)$ $(k = 0, 1, \ldots, l)$, the lower plot presents the relative concentrations of individual sequences, $\bar{x}_j(p)$ $(j = 0, 1, 2, \ldots, 2^l - 1)$. Choice of parameters: $l = 50$, $f_0 = 10$, $f_n = 1$, and $\varphi = 1/(10^{1/50})$.

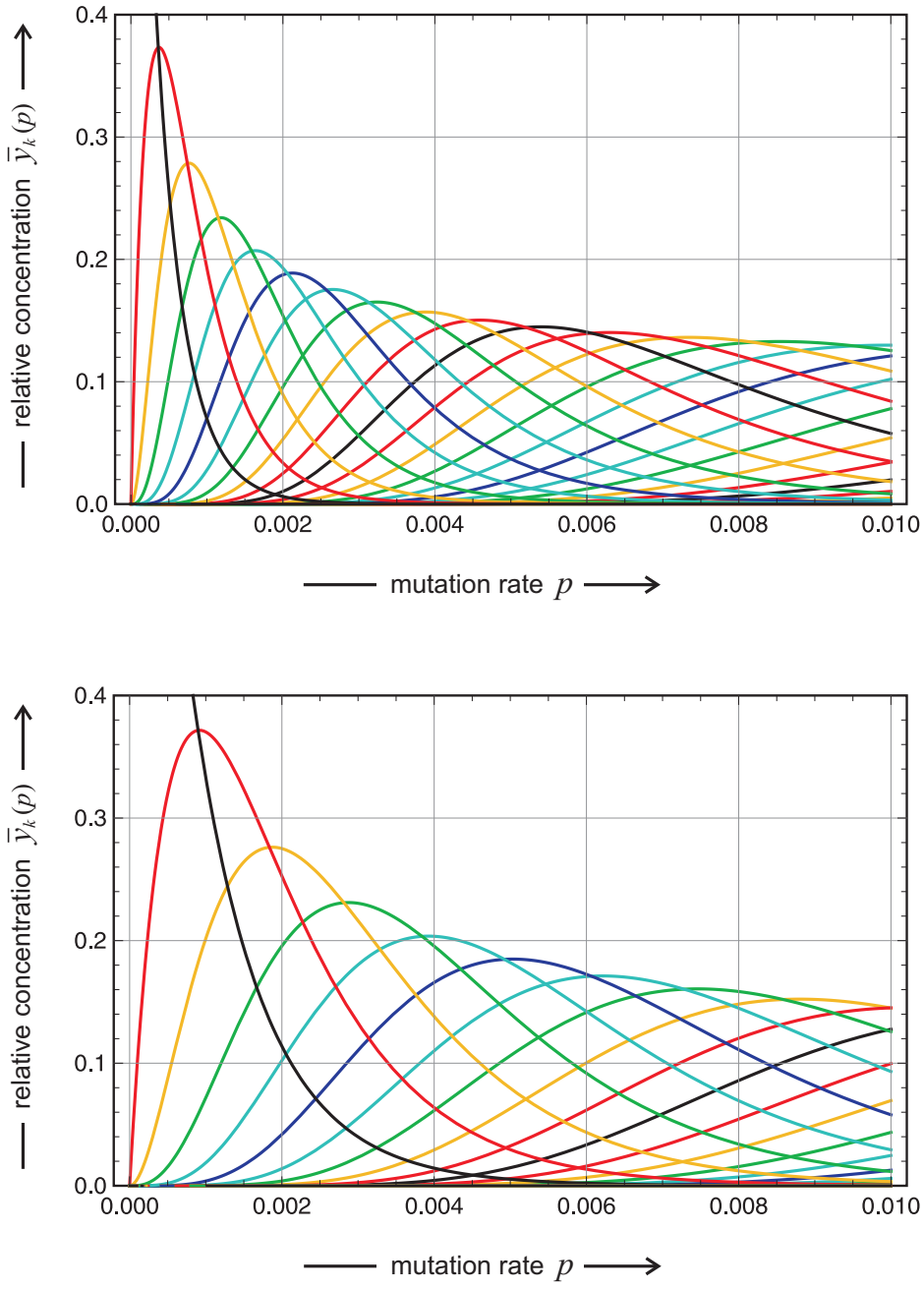


Figure 4.17: **Comparison of quasispecies on the additive and the multiplicative landscape**. The plots show the dependence of the quasispecies from binary sequences on the mutation rate parameter $\overline{\Upsilon}(p)$. The upper plot shows the relative concentrations of entire mutant classes, $\bar{y}_k$ $(k = 0, 1, \ldots, l)$, on the additive landscapes (3.4a), and the lower plot presents the plot for the multiplicative landscapes (3.4b). Choice of parameters: $l = 50$, $f_0 = 10$, $f_n = 1$, $\theta = 9/50$, and $\varphi = 1/(10^{1/50})$.

the transition is more difficult to locate but the formation and broadening of the plateau representing the domain of the approximate uniform distribution is easily recognized, and as the plateau broadens the decay regions of the master sequence (lhs) or the master pair (rhs) are compressed. Fig. 4.15 finally shows enlargements of the regimes of direct and complementary replication. In the enlargement the error threshold in direct replication is still broad compared to the $l = 50$ case (Fig. 4.8). In complementary replication we observe that the curves for $\mathbf{X}_0$ and $\mathbf{X}_{31}$ approach each other as $f_0 - f_n$ becomes smaller: The values for $\bar{x}_{31}/\bar{x}_0$ are $\sqrt{10} \approx 3.162$, 2, and $\sqrt{1.1} \approx 1.049$, respectively. In is interesting to observe that in the enlargement for $f_0 = 1.1$ the curves for the complementary pairs in the one- and four-error classes two as well as in the two- and three-error classes are so close together that they can no more be resolved for $p < 1$.

Two among the model landscapes presented in Equ. (3.4), the additive or linear and the multiplicative or exponential landscape show no error threshold in the sense of the single peak landscape. These two landscapes are very similar and it is sufficient therefore to discuss only one of them and to show one comparison here, and we choose the multiplicative landscape for the purpose of illustration because it is used more frequently in population genetics. Although the concentration of the master sequence decays very fast on the multiplicative landscape (Fig. 4.16) no error threshold phenomenon is observed. Instead the curves for the individual error classes look plaited around themselves. Looking more carefully, however, shows that the each curve $\bar{y}_k$ passes a single maximum only ($k > l/2$) or no maximum at all ($k \leq l/2$) before is converges to $\bar{y}_k(\frac{1}{2}) = (\frac{1}{2})^l \binom{l}{k}$. This appearance of the plot for the classes is a result of the binomial factors, and the plot the individual concentrations $\bar{x}_j(p)$ reflects simply the lower frequencies for sequences with larger Hamming distance from the master sequence, $d_{\mathrm{H}}(\mathbf{X}_j, \mathbf{X}_m)$. It is worth mentioning that application of a smaller fitness difference shifts all curves closer to p_0 and compresses them without, however, changing the overall appearance.

In qualitative terms the additive landscape gives rise to the same curves as the multiplicative landscape (Fig. 4.17). The quantitative comparison for

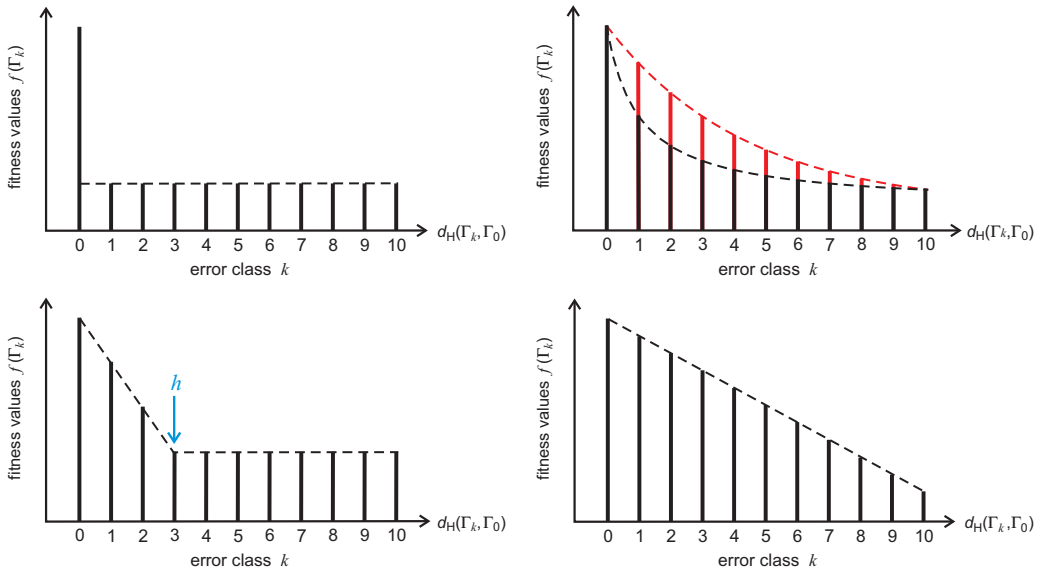


Figure 4.18: **Some examples of model fitness landscapes**. The figure shows five model landscapes with identical fitness values for all sequences in a given error class: (i) the single peak landscape (upper left drawing), (ii) the hyperbolic landscape (upper right drawing, black curve), (iii) the step-linear landscape (lower left drawing), (iv) the multiplicative landscape (upper right drawing, red curve), and (v) the additive or linear landscape (lower right drawing). Mathematical expressions are given in the text.

the same highest and lowest fitness values shows that the decay of the master sequence is steeper and and the curves are more compressed on the additive landscapes. The interpretation is straightforward: The highest fitness value except the master is larger on the additive landscape, $f_1 = f_0 - (f_0 - f_n)/l = 9.82$ versus $f_1 = f_0(f_n/f_0)^{1/l} \approx 9.55$ on the multiplicative landscape.

There are, of course, many other possible simple landscapes that show, in essence, one these two scenarios or combinations of both [250]. In the next subsection we shall compare several simple landscapes and analyze the error threshold in more detail. In particular, we shall show that the threshold as it occurs on the single peak landscape is a superposition of three phenomena that are separable through appropriate choice of the distribution of fitness values.

4.3.6 Error thresholds on "simple" model landscapes

The fact that existence and form of an error threshold depend on the nature of the fitness landscape has been pointed out already fifteen years ago [308]. As has been discussed in the previous subsection 4.3.5 the additive and the multiplicative landscape – the two types of landscapes that are commonly used in population genetics – don't show error thresholds at all. In addition to these two landscapes and the single peak landscape we consider two further examples of simple model landscapes, the hyperbolic landscape and the step-linear landscape. All these *simple landscapes* are characterized by identical fitness values for all members of the same mutant class. In particular, the fitness values are of the form:

(i) the additive or linear landscape
$f(\mathbf{Y}_k) = f_k = f_0 - (f_0 - f)\, k/l\,;\ k = 0, 1, \ldots, l\ ,$

(ii) the multiplicative landscape
$f(\mathbf{Y}_k) = f_k = f_0 \left(\frac{f}{f_0}\right)^{k/l}\ ;\ k = 0, 1, \ldots, l\ ,$

(iii) the hyperbolic landscape
$f(\mathbf{Y}_k) = f_k = f_0 - (f_0 - f)\,(\frac{l+1}{l})\,(\frac{k}{k+1})\,;\ k = 0, 1, \ldots, l$, and

(iv) the step-linear landscape
$$f(\mathbf{Y}_k) = f_k = \begin{cases} f_0 - (f_0 - f)\, k/h & \text{if } \ k = 0, 1, \ldots, h-1\,, \\ f & \text{if } \ k = h, \ldots, l\ . \end{cases}$$

In order to be able to compare the different landscapes the values of f were chosen such that all landscapes are characterized by the same superiority of the master sequence: $\sigma_m = \sigma_0 = f_0 \,/\, \overline{f}_{-0}$ with $\overline{f}_{-0} = \sum_{i=1}^{n} y_i f_i \,/\, (1 - y_0)$. Since the distribution of concentrations is not known *a priori* we have to make an assumption. As shown in subsection 4.3.3 the range of (approximate) validity of the uniform distribution extends in the direction of decreasing mutation rates from the point $p = \tilde{p}$ far down to the error threshold for and hence, the assumption of the uniform distribution in the calculation of f is

Table 4.3: **Concentration level crossing near the error threshold.** The decline of the master class, $\bar{y}_0 = \bar{x}_0$, at p-values below the error threshold $p_{\rm cr}$ is illustrated by means of the points $p_{(1/M)}$ where $\bar{y}_0(p)$ crosses the level $1/M$ for the three fitness landscapes that sustain error thresholds. Parameters: $l = 100$, $f_0 = 10$, and $\overline{f}_{-0} = 1$.

Landscape	Level crossing			Error threshold
	$p_{(1/100)}$	$p_{(1/1000)}$	$p_{(1/10000)}$	$p_{\rm cr}$
Single-peak	0.02198	0.02274	0.02282	0.02277
Hyperbolic	0.01450	0.01810	0.02036	0.02277
Step-linear	0.01067	0.01774	0.02330	0.02277

well justified,

$$f = \left(\overline{f}_{-0}\,(2^l - 1) - f_0\,(2^{l-1} - 1)\right) \Big/ 2^{l-1}\,, \tag{4.34a}$$

$$f = \left(\left(\overline{f}_{-0}\,(2^l - 1) + f_0\right)^{1/l} - f_0^{1/l}\right)^l\,, \tag{4.34b}$$

$$f = \left(\overline{f}_{-0}\,l\,(2^l - 1) - f_0\,(2^l - l + 1)\right) \Big/ \left(2^l(l-1) + 1\right)\,, \tag{4.34c}$$

$$f = \frac{\overline{f}_{-0}\,(2^l - 1) - f_0\left(\sum_{k=0}^{h-1}\binom{l}{k}\frac{h-k}{h} - 1\right)}{\sum_{k=0}^{h-1}\binom{l}{k}\frac{k}{h} + \sum_{k=h}^{l}\binom{l}{k}}\,. \tag{4.34d}$$

In Fig. 4.19 and 4.20 the solution curves $\bar{y}_k(p)$ are compared for the three landscapes showing error thresholds, single-peak, hyperbolic and step-linear. The superiority was adjusted to $\sigma_0 = 10$ be means of equation (4.34).

The additive and the multiplicative landscape do not sustain sharp transitions but show a gradual transformation of the master dominated quasispecies to the uniform distribution being the exact solution of the mutation-selection equation (4.9) at $p = \tilde{p} = \kappa^{-l}$, which has been discussed in the previous subsection 4.3.5. Here we shall concentrate on landscapes that give rise to sharp transitions as we observed it previously on the single peak

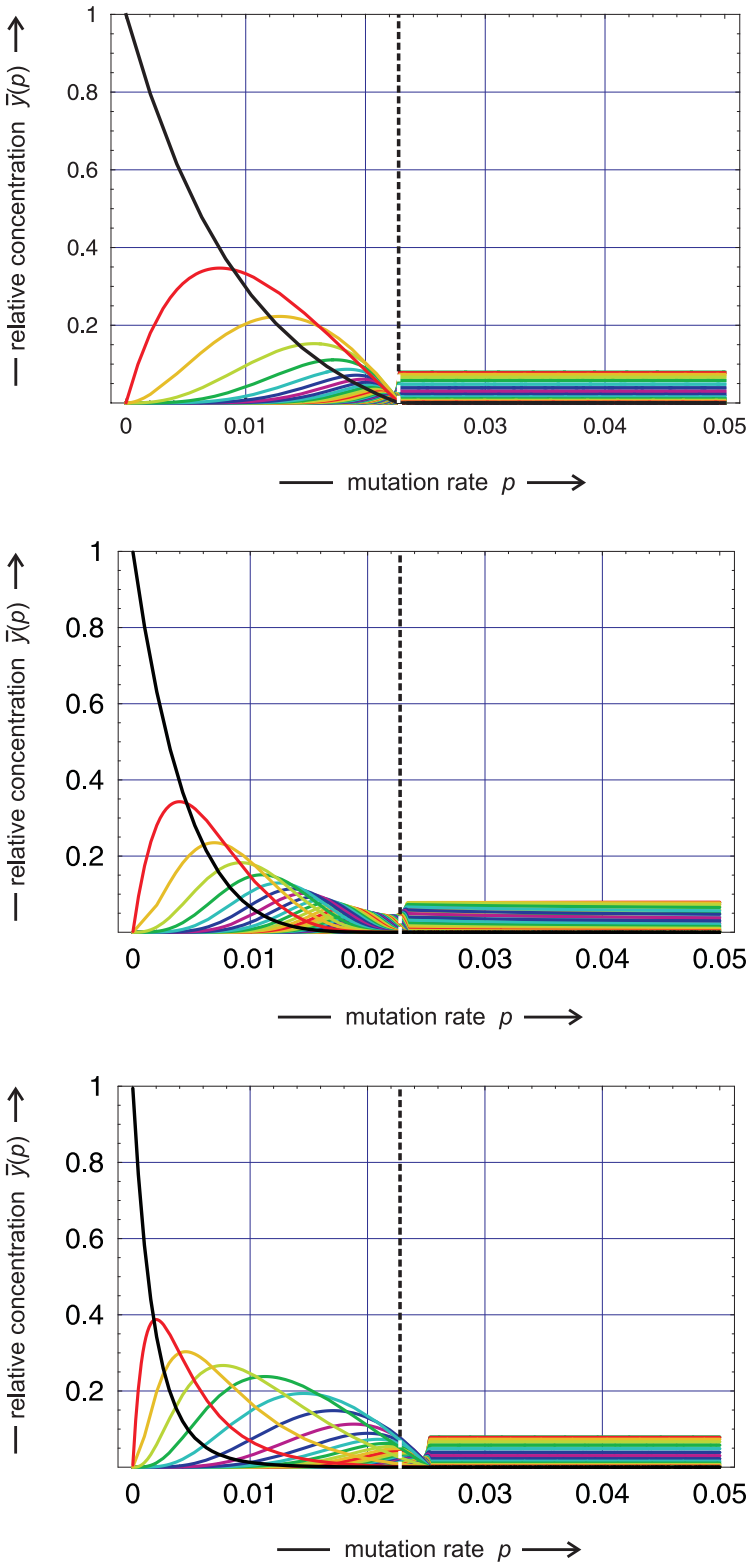


Figure 4.19: **Error thresholds on different model landscapes.** The figures show stationary concentrations of mutant classes as functions of the error rate, $\bar{y}_k(p)$, for sequences of chain length $l = 100$ with $f_0 = 10$ and $\overline{f}_{-0} = 1$ on three different model landscapes: the single peak landscape (upper part, $f = 1$), the hyperbolic landscape (middle part, $f = 10/11$), and the step-linear landscape (lower part, $f = 1$). The dashed line indicates the value of the error threshold

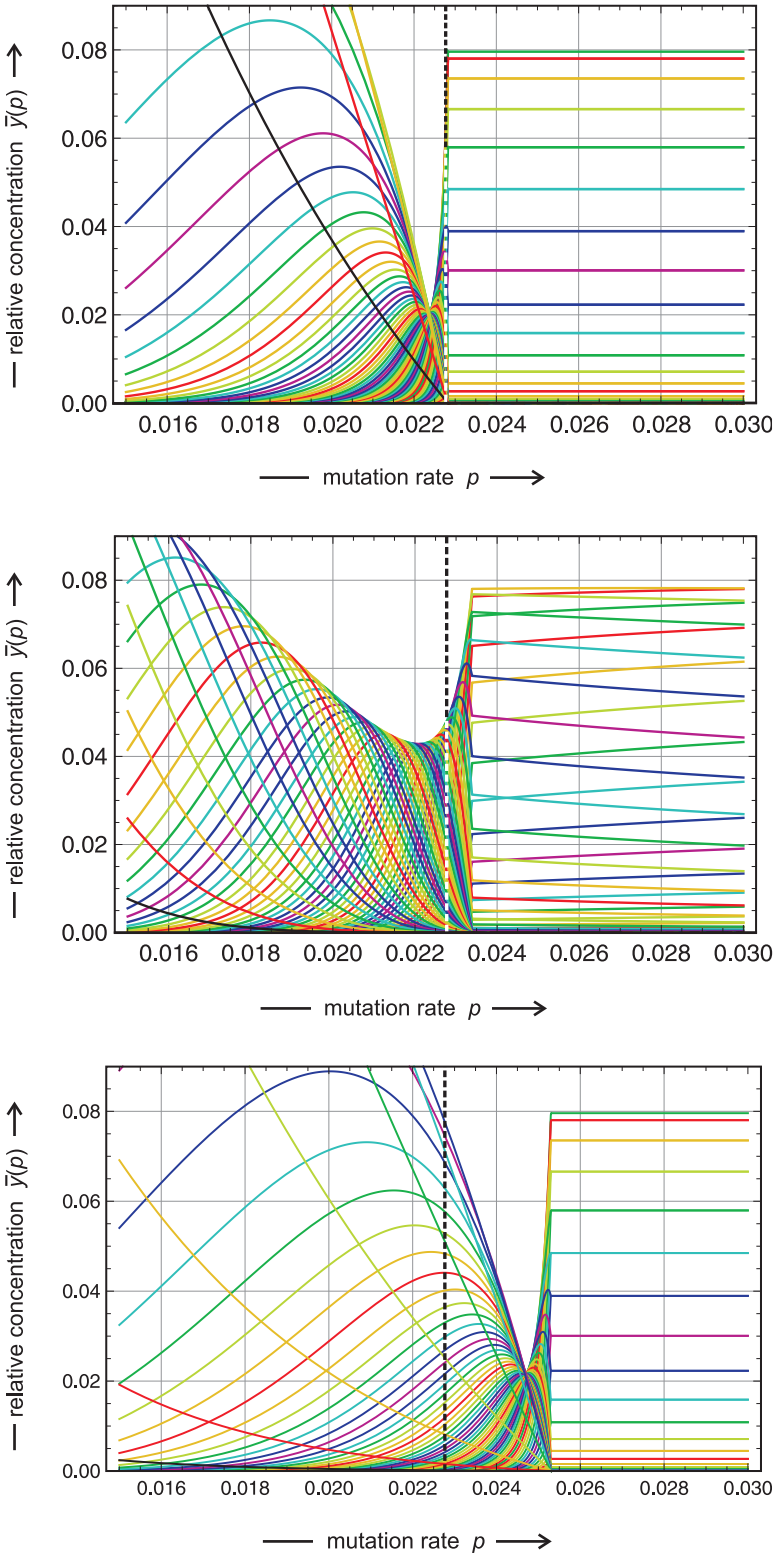


Figure 4.20: **Error thresholds on different model landscapes.** The three figures are enlargements of the plots from in Fig. 4.19. Stationary concentrations of mutant classes, $\bar{y}_k(p)$, are shown for the single peak landscape (upper part), the hyperbolic landscape (middle part), and the step-linear landscape (lower part; see the caption Fig. 4.19 for details).

landscape. The three examples chosen differ in the way the fitness function changes with k the index of the mutant class (Γ_k): (i) the "L" shaped single peak landscape, (ii) the hyperbolic landscape that shows fast steep but at the same gradual decay of the fitness with k, and (iii) the step-linear landscape that combines features of the linear and the single peak landscape – linear decay of fitness values and completely flat part.

As already shown in Fig. 4.8 and analyzed in the proof for the occurrence of the threshold (Tab. 4.1) the calculated value for p_{cr} coincides perfectly with the position of the transition on the single-peak landscape, and the p-values for concentration level crossing lie close together and near p_{cr} (see table 4.3) indicating a rather steep decrease of $\bar{y}_0$ in the range on the left-hand side of the transition to the uniform distribution. Comparison with the other two landscapes shows that the error threshold phenomenon can be separated into three different features, which happen to coincide on the single peak landscape: (i) fast decay of the concentration of the master sequence $\bar{x}_0(p)$ in the range $0 \leq p \leq p_{\mathrm{cr}}$, (ii) a sharp transition to another sequence distribution and (iii) an extension of the solution at point $p = \tilde{p}$ towards smaller mutation rates that gives rise to a broad domain $p_{\mathrm{cr}} < p < \tilde{p}$ within which the quasispecies is very close to the uniform distribution.

On the hyperbolic landscape, the actual transition occurs slightly above the error threshold of the single-peak landscape, the decrease of $\bar{y}_0$ is flatter, and the transition also sharp does not result in the uniform distribution. Instead we observe a mutant distribution above the error threshold, which changes slightly with p.The interpretation is straightforward: A flat part of the landscape in required for the expansion of the unform distribution and such a flat part does not exist on the hyperbolic landscape, and therefore we observe a gradually changing flat distribution. On the step-linear landscape, eventually, the curve of $\bar{y}_0$ is even flatter, the transition is shifted further to higher p-values, but the transition leads to the uniform distribution as in the single-peak case. Knowing the behavior of the quasispecies on the single-peak and the linear landscape an interpretation of the observed plots for the step-linear landscape is straightforward: In the range of small Hamming distances from the master sequence the fitness landscape has the same

shape as the linear landscape and for small mutation rates the quasispecies is dominated by sequences, which are near the master in sequence space, at higher mutation rates p sequences that are further away from the master gain importance, and indeed we observe a similarity of the quasispecies with that on the linear landscape at small p-values whereas an error threshold and the uniform distribution beyond it are observed at higher mutation rates p. In the step-linear landscape the position of the step, h can be varied as well. For the parameters $f_0 = 10$ and $f = 1$ we observe error thresholds in the range $0 \leq h \leq 35$, at higher h-values it becomes softer and eventually around $h = 45$ it has completely disappeared.[21] A useful indicator for the existence of an error threshold is the upper envelope of all individual curves $\bar{y}_k(p)$: The absence of a threshold leads to a monotonous decrease off the envelope (Figs. 4.16 and 4.17) whereas an error threshold manifests itself in a pronounced minimum of the envelope just below p_{cr} (Fig. 4.20).

The fact that the behavior of quasispecies depends strongly on the nature of the fitness landscape is not surprising. Fitness values after all play the same role as rate parameters in chemical kinetics and the behavior of a system can be changed completely by a different choice of rate parameters. The most relevant but also most difficult question concerns the relation between rate parameters and observed stationary distributions: Can we predict the quasispecies from a knowledge of the fitness landscape? Or the even more difficult inverse problem [77]: Does the observed behavior of the quasispecies allow for conclusions about the distribution of fitness values? A few regularities were recognized already from observations on the simple model landscapes: (i) steep decay of the master concentration, $\bar{y}_0(p)$ may occur without the appearance of a sharp transition, (ii) a sharp transition may occur on fitness landscapes with gradually changing fitness values provided the decay of $f(\mathbf{Y}_k)$ with k is sufficiently steep, (iii) a sharp transition may occur without leading to the uniform distribution, and (iv) the appearance of the uniform distribution at p_{cr}-values lower than $\tilde{p}$ requires a flat part of

[21] Like in physics we distinguish hard and soft transitions. A hard transition is confined to a very narrow range of the order parameter – here the error rate p – and becomes steeper and steeper as the system grows to infinity.

the fitness landscapes in the sense that fitness values of neighboring classes are the same.

5. Fitness landscapes and evolutionary dynamics

Sewall Wright's metaphor of a fitness landscape has been considered far off reality for more than forty years. Growing knowledge on biopolymer structure and function, in particular structural data from proteins and RNA molecules, led to first speculations on the nature of sequence structure mappings. The notion of sequence space was suggested [65, 201] for the support in mappings and landscapes based on combinatorial diversity. Relatedness in sequence space was derived from the mutational distance that is commonly identified with the Hamming distance between sequences, $d_{\mathrm{H}}(\mathbf{X}_j, \mathbf{X}_i) = d_{ji} = d_{ij}$, which induces a metric and hence is symmetric with respect to the aligned sequences. Still the tools for systematic searches in sequence space were not yet available and it needed the new sequencing methods as well as high-throughput techniques before progress in the empirical determination of fitness landscapes became possible. Much later, mainly based on experimentally as well as theoretically derived properties of biomolecules the notion of a *structure space* or *shape space*[1] has been created [98, 238, 252] that corresponds to a *phenotype space* in evolution research.

In this chapter we present at first the landscape concept from the point of view of a structural biologist. Fitness and other properties can be derived from molecular functions, which are thought to be essentially determined by molecular structures. RNA secondary structures are used as a model system, which is sufficiently simple to be analyzed by rigorous mathematics but encapsulates at the same time the essential features of more complex sequence-structures mappings. Two features are characteristic for landscapes derived from biopolymers, in particular RNA molecules, are: (i) *ruggedness* – pairs of sequences situated nearby in sequence space, i.e., having Hamming distance $d_{\mathrm{H}} = 1$, may give rise to very similar or entirely different structures

[1] The notion of shape space is commonly used also in mechanical engineering for the complete set of shapes that can be assembled from a few elementary objects.

and properties – and (ii) *neutrality* – two or more sequences may have identical structures and properties,[2] and these neutral sequences form *neutral networks* in sequence space, which are the preimages of the structures (see Fig. 5.2).

In the second part we introduce model fitness landscapes that reproduce the basic characteristics of functions derived from biopolymer structures and study evolutionary dynamics on them. The dynamics on these "realistic" fitness landscapes reflects several features that were observed on simple fitness landscapes already (subsection 4.3.5) like the existence of error thresholds but reveals also new phenomena like phase-transition like conversions between different quasispecies. The third part is dealing with neutrality and deterministic evolutionary dynamics in presence of two or more neutral sequences. This section is complementary to the chapter on stochasticity (chapter 8).

5.1 RNA landscapes

The majority of data on the relation between sequences and molecular properties comes from structural biology of biopolymers, in particular RNA and protein. As said RNA secondary structures are chosen here because they provide a simple and mathematically accessible example of a *realistic* mapping of biopolymer sequences onto structures [248, 252]. The RNA model is commonly restricted to the assignment of a single structure to every sequence but the explicit consideration of suboptimal conformations is possible as well [248] (see Fig. 5.7) and will be used here to illustrate more complex functions of RNA molecules, for example switches controlling metabolism. Neutrality with respect to structure formation implies that several sequences fold into the same structure or, in other words, the RNA sequence-structure mapping is not invertible.

Originating from the application of quantum mechanics to molecular motions the Born-Oppenheimer approximation gave rise to molecular *hypersurfaces* upon which nuclear motion takes place. Meanwhile the landscape

[2] *Identical* in the context of neutrality does not mean identical in strict mathematical sense but indistinguishable for the experimental setup or for natural selection [174, 227].

concept became also an integral part of biophysics and in other areas of physics and chemistry. In particular, conformational landscapes of biopolymers have been and are successfully applied to the folding problem of proteins [229, 314]. Nucleic acid structures, in particular RNA in the simplified form of secondary structures, turned out to provide a sufficiently simple model for the study of basic features of sequence-structure mappings [238, 239]. What renders nucleic acids accessible to mathematical analysis is the straightforward and unique base pairing logic for nucleic acids: In DNA **A** pairs with **T** and **C** pairs with **G** providing thereby the basis of replication and reproduction. Base pairing by the same token dominates intramolecular interactions in RNA and accounts for the major fraction of the free energy of folding. Base pairing in RNA, however, is slightly relaxed: As in DNA have the Watson-Crick pairs $\mathbf{A} = \mathbf{U}$ and $\mathbf{G} \equiv \mathbf{C}$ but the wobble pairs $\mathbf{G}-\mathbf{U}$ are accepted as well in secondary structures of single stranded RNA. Base pairing logic, for example, allows for the application of combinatorics in counting of structures with predefined structural features or properties [146, 300]. In addition, efficient algorithms based on dynamic programming and using empirically determined parameter sets are available for RNA structure prediction [145, 326, 328].

5.1.1 The paradigm of structural biology

In structural biology the relation between biopolymer sequences and functions is conventionally spit into two parts: (i) the mapping of sequences into structures and (ii) the prediction or assignment of function for known structures (Fig. 5.1). If function is encapsulated in a scalar quantity, for example in a reaction rate parameter or a fitness value, the second mapping corresponds to a landscape with structure space as support:

$$\mathbf{X} \longrightarrow \mathbf{S} = \Phi(\mathbf{X}) \longrightarrow f = \Psi(\mathbf{S}) \; . \tag{5.1}$$

The function itself gives rise to the dynamics of a process involving a population of sequences or genotypes. The underlying concept is based on the assumption that structures can be calculated from sequences either directly or by means of an algorithm. Function manifests itself in the structure and

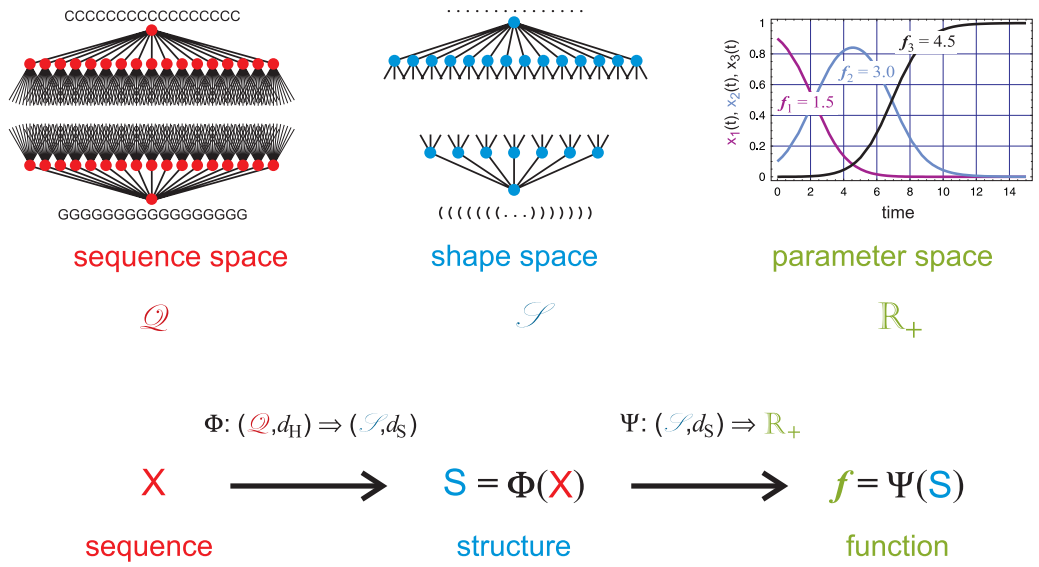


Figure 5.1: **The paradigm of structural biology.** The relations between sequences, structures, and functions involve three spaces: (i) the sequence space $\mathcal{Q}$ with the Hamming distance d_H as metric, (i) the shape space $\mathcal{S}$ with a structure distance d_S as metric, and (iii) the parameter space $\mathbb{R}_+^m$ for m parameters. Properties and functions are viewed as the result of two consecutive mappings: Φ maps sequences into structures, Ψ assigns parameters to structures and thereby determines molecular function (The insert shows fitness values f_k and selection as as example). The sequence space $\mathcal{Q}$ has a remarkable symmetry: All sequences are equivalent in the sense that the occupy topologically identical positions having the same number of nearest, next nearest, etc., neighbors linked in the same way (Examples of a sequence spaces are shown in Figs. 3.3, 3.4, and 4.7). The shape space $\mathcal{S}$ refers here to RNA secondary structures, which can be uniquely represented by strings containing parentheses for base pairs and dots for single stranded nucleotides (see Fig. 5.3 for an explanation). The elements of shape space can be classified by the number of base pairs and then there is a unique *smallest* element, the open chain, and depending on l one *largest* element for odd l or two largest elements for even l – one with the unpaired nucleobase on the 5'-end and one with it on the 3'-end. Parameter space in chemical kinetics is commonly multi-dimensional and the elements are rate parameters, commonly nonnegative real numbers $f_k \in \mathbb{R}_+$; $k = 1, \ldots$, or equilibrium properties like binding parameters.

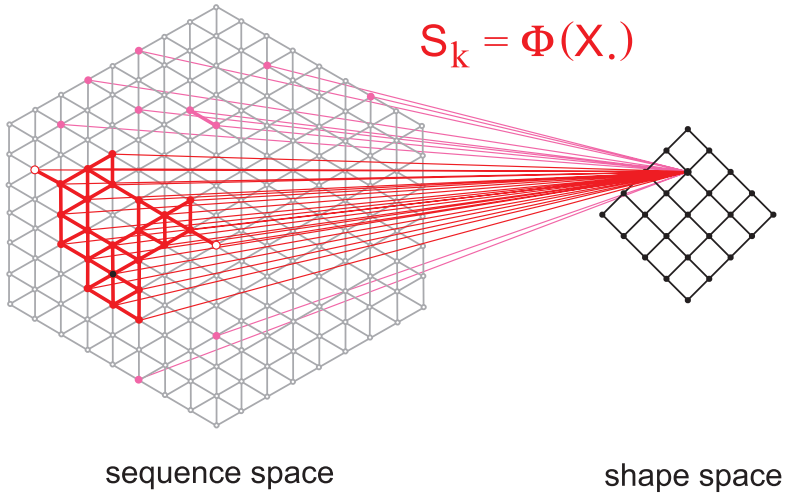


Figure 5.2: **A sketch of the mapping of RNA sequences onto secondary structures.** The points of sequence space (here 183 on a planar hexagonal lattice) are mapped onto points in shape space (here 25 on a square lattice) and, inevitably, the mapping is many to one. All sequences $\mathbf{X}_\cdot$ folding the same mfe structure form a neutral set, which in mathematical terms is the preimage of $\mathbf{S}_k$ in sequence space. Connecting nearest neighbors of this set – these are pairs of sequences with Hamming distance $d_{\mathrm{H}} = 1$ – yields the neutral network of the structure, $\mathcal{G}_k$. The network in sequence space consists of a giant component (red) and several other small components (pink). On the network the stability against point mutations varies from $\hat{\lambda} = 1/6$ (white points) to $\hat{\lambda} = 6/6 = 1$ (black point). We remark that the two-dimensional representations of sequence are used here only for the purpose of illustration. In reality, both spaces are high-dimensional – the sequence space of binary sequences $\mathcal{Q}_l^{(2)}$, for example, is a hypercube of dimension n and that of natural four-letter sequences $\mathcal{Q}_l^{(4)}$ an object in $3n$ dimensional space.

should be predictable therefore. As it turned out after some spectacular successes in the early days (see, e.g., [302]) both steps are feasible but the mappings are highly complex and not fully understood yet. The alternative way to determine parameters consists of an inversion of conventional kinetics: The measured time dependence of concentrations is the input and parameters are fit to the data either by trial and error or systematically by means of inverse methods involving regularization techniques [77].

Counting sequences and structures reveals the ultimate basis for neutrality. In the natural **AUGC** alphabet we are dealing with $\kappa^l = 4^l$ RNA

sequences whereas the number of acceptable secondary structures[3] that has been determined by combinatorics [146, 300]:

$$\left|\mathcal{S}_{\text{lim}}^{(3,2)}\right|(l) = 1.4848 \times l^{-3/2}(1.84892)^l \ . \tag{5.2}$$

The formula is asymptotically correct for long sequences as indicated by the limit. Insertion of moderately long or even small chain lengths l into Equ. 5.2 shows that we are always dealing with orders of magnitude more sequences that structures and therefore neutrality with respect to structure and the formation of neutral network for common structures is inevitable [247].

The conventional problem in structural biology is to find the structures into which a sequences folds under predefined conditions [327, 328]. Solutions to the inverse problem, finding a sequence that folds into a given structure are important for the design of molecules in synthetic biology. An inverse folding algorithm has been developed for RNA secondary structures [145] (for a recent variant see [2]) and turned out to be a very useful tool for studying sequence-structure mappings. In particular, these mappings are not invertible: Many sequences fold into the same structure, and the notion of *neutral network* has been created for the graph representing the preimage of a given structure in sequence space [252]:

$$\Phi(\mathbf{X}_j) \ = \ \mathbf{S}_k \ \ \Longrightarrow \ \ G_k \ = \ \Phi^{-1}(\mathbf{S}_k) \ \equiv \ \{\mathbf{X}_j | \Phi(\mathbf{X}_j \ = \ \mathbf{S}_k\} \ , \tag{5.3}$$

The neutral set G_k in converted into the *neutral network* $\mathcal{G}_k$ through connecting all pairs of sequences with Hamming distance $d_{\text{H}} = 1$. The definition of the neutral networks in a way restores uniqueness of the mapping: Every structure $\mathbf{S}_k$ has a uniquely defined neutral network $\mathcal{G}_k$ in sequence space. Every sequence $\mathbf{X}_j$ belongs to one and only one neutral network. Neutral networks are characterized by a (mean) degree of neutrality

$$\bar{\lambda}_k \ = \ \frac{\sum_{j|\mathbf{X}_j \in G_k} \lambda_j^{(k)}}{|G_k|} \ , \tag{5.4}$$

[3]Acceptable means here that hairpin loops of lengths $n_{\text{hp}} = 1$ or 2 are excluded for stereochemical reasons and stacks of length $n_{\text{st}} = 1$, i.e. isolated base pairs, are not considered for poor energetics (see Fig. 5.4).

wherein $\lambda_j^{(k)}$ is the local fraction of neutral nearest neighbors of sequence $\mathbf{X}_j$ in sequence space (An example is sketched in Fig.5.2). Without knowing special features of neutral networks derived from RNA secondary structures the application of *random graph theory* [78, 79] to neutral networks is obvious. One (statistical) result of the theory dealing with the connectedness of the graph applies straightforwardly and relates it with the degree of neutrality $\bar{\lambda}_k$: Neutral networks $\mathcal{G}_k$ with a degree of neutrality above a critical value, $\bar{\lambda}_k > \lambda_{\text{cr}}$, are connected whereas networks with lower degree of neutrality, $\bar{\lambda}_k < \lambda_{\text{cr}}$, are partitioned into components with one particularly large and several small components. The large component is commonly characterized as *giant component*. The critical degree of neutrality depends only on the size of the nucleobase alphabet: $\lambda_{\text{cr}} = 1 - \kappa^{-1/(\kappa-1)}$ leading to $\lambda_{\text{cr}} = 0.5$ for $\kappa = 2$ and to $\lambda_{\text{cr}} = 0.370$ for $\kappa = 4$. Because of the non-invertibility of the mapping distance relations in structure space are different from those in sequence space and they are more complex. The notion of distance is replaced by a concept on nearness [102] of neutral networks. Nearness, however, does not fulfill the properties of a metric as careful mathematical analysis reveals but leads to a pretopology in phenotype space [270]. Fig.5.2 shows a sketch of a typical neutral network in a 'two-dimensional' sequence space. The network consists of several components with one *giant component* being much larger than the others. Random graph theory was found to represent a proper reference also for the neutral networks based on RNA secondary structures in the sense that deviations from the idealized node distribution can be interpreted by structure based nonhomogeneous distributions of sequences in sequence space [127]. Some structures with special features like, for example, stack with free ends on both sides may show two or four giant components of equal size or three giant components with a size distribution 1:2:1.

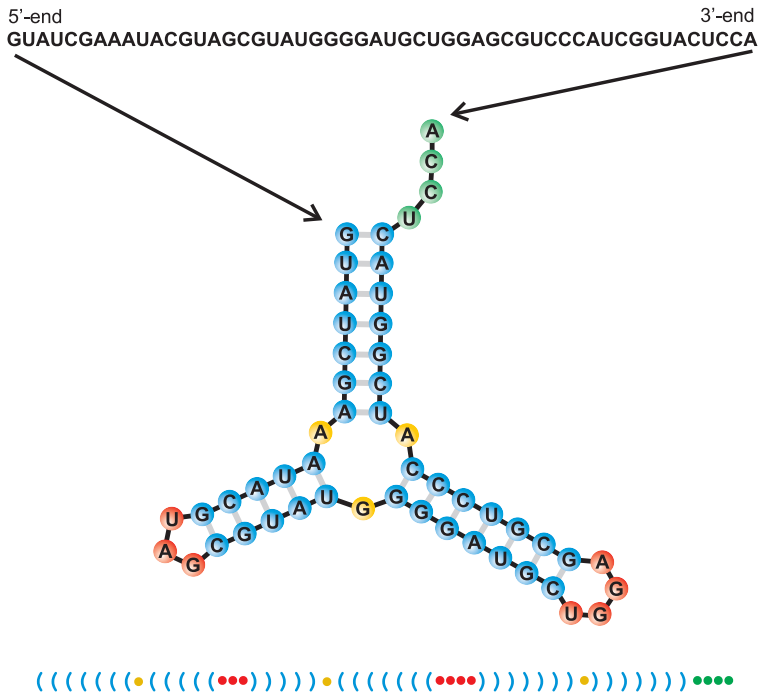


Figure 5.3: **Folding of RNA sequences into secondary structures.** A sequence of chain lengths $l = 52$ is converted into its most stable or *minimum free energy* (mfe) structure – represented here by the conventional graph – be means of a folding algorithm. The most popular algorithms are use the technique of dynamic programming to find the structure of lowest free energy [145, 326, 328]. The parameter values, enthalpies and entropies of formation for structural elements and substructures (Fig. 5.4) are determined empirically [199, 200]. The string below the graph is an equivalent representation of the secondary structure: Base pairs are represented by parentheses and single stranded nucleotides by dots. Color is used here only to facilitate the assignment of graph substructures to stretches on the structure string. It is not required to make the assignment of the graph to the string and vice versa unique.

5.1.2 RNA secondary structures

An RNA secondary structure is a listing of base pairs that is conventionally cast into a graph.[4] The secondary structure graph (Fig. 5.3) is obtained

[4]It is important to note that a graph does not represent a structure. It defines only neighborhood relations. In case of the RNA secondary structures these are (i) the neighborhood in the RNA backbone corresponding to the neighborhood in the sequence and (ii) the neighborhood in the base pairs.

by means of efficient algorithms using dynamic programming [145, 326, 328]. The basic idea is partitioning of the secondary structure into structural elements, double stranded base pairs and single stranded loops and external elements, which are assumed to contribute independently to the total free energy of the molecule. An illustration of individual substructures and their combination to RNA structures is shown in Fig. 5.4. The desirable uniqueness of folding requires a specification of the conditions under which the process takes place. Commonly, the criterium of folding is finding the thermodynamically most stable structure called *minimum free energy* (mfe) structure, which makes the implicit assumption that the process may take infinitely long time. Another important condition is folding of the growing chain during RNA synthesis. For long RNA molecules melting and refolding may take very long time (see subsection 5.1.3) and then the result obtained by folding on-the-fly is metastable and represents a conformation that is different from the mfe-structure. In addition, folding kinetics may prefer conformations that are different from the mfe structure.

The two characteristic features of landscapes derived from biopolymer structures, which were mentioned initially – ruggedness and neutrality, become immediately evident through an analysis of RNA structures. This simultaneous appearance of ruggedness and neutrality is illustrated most easily by means of RNA secondary structures, which are defined in terms of Watson-Crick and $\mathsf{G}-\mathsf{U}$ base pairs: Exchange of one nucleobase in a base pair, e.g., $\mathsf{C} \to \mathsf{G}$ in $\mathsf{G} \equiv \mathsf{C}$, may open the base pair, destroy a stack, and eventually lead to an entirely different structure with different properties, or leave the structure unchanged, e.g., $\mathsf{A} \to \mathsf{G}$ in $\mathsf{A} = \mathsf{U}$. Neutrality is equally well demonstrated: Exchanging both bases in a base pair may leave structure and (most) properties unchanged, $\mathsf{G} \equiv \mathsf{C} \ \to \ \mathsf{C} \equiv \mathsf{G}$ may serve as an example. Evolutionary dynamics is clearly influenced by the shape of fitness landscapes and the interplay of the two characteristic features was found to be essential for the success of evolutionary searches [101, 102, 156].

In order to illustrate the typical form of the local environment in a biopolymer landscape we choose a small RNA of chain length $l = 17$ with

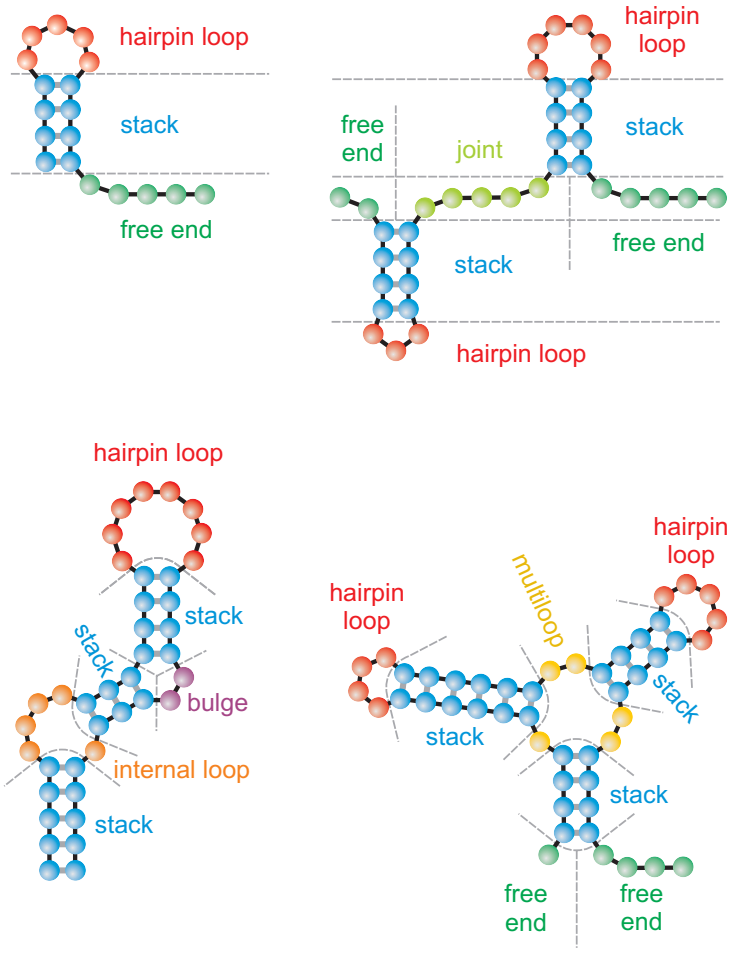


Figure 5.4: **Elements of RNA secondary structures.** Three classes of structural elements are distinguished: (i) stacks (indicated by nucleotides in blue color), (ii) loops, and (iii) external elements being joints (chartreuse) and free ends (green). Loops fall into several subclasses: Hairpin loops (red) have one base pair, called the closing pair, in the loop. Bulges (violet) and internal loops (orange) have two closing pairs, and loops with three or more closing pairs (yellow) are called multiloops. The number of closing pairs is the *degree* of the loop.

the sequence $\mathbf{X}_0 \equiv$**AGCUUACUUAGUGCGCU** as example.[5] At 0 °C the sequence forms the minimum free energy structure $\mathbf{S}_0 = \Phi(\mathbf{X}_0)$, which consists of an hairpin with six base pairs and an internal loop that separates two

[5] All polynucleotide sequences are written from the 5'-end at the lhs to the 3'-end on the rhs.

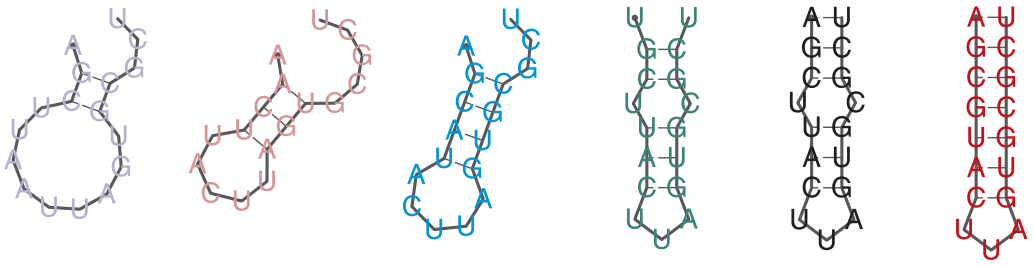


Figure 5.5: **Selected RNA structures.** Shown are examples of RNA structures in the one-error neighborhood of sequence $\mathbf{X}_0 \equiv$(**AGCUUACUUAGUGCGCU**) ordered by the numbers of base pairs. In total, the $3 \times 17 = 51$ sequences form two different structures with two base pairs (1,8; the numbers in parentheses refer to the occurrence of the individual structures), four structures with three base pairs (1,1,2,3), three structure with four base pairs (1,2,3), four structures with five base pairs (1,1,3,4), two structures with six base pairs (2,15), and one structure with seven base pairs (3). The three structures on the rhs have a common folding pattern and differ only by closing and opening of a base pair: (i) the two bases in the internal loop and (ii) the outermost base pair.

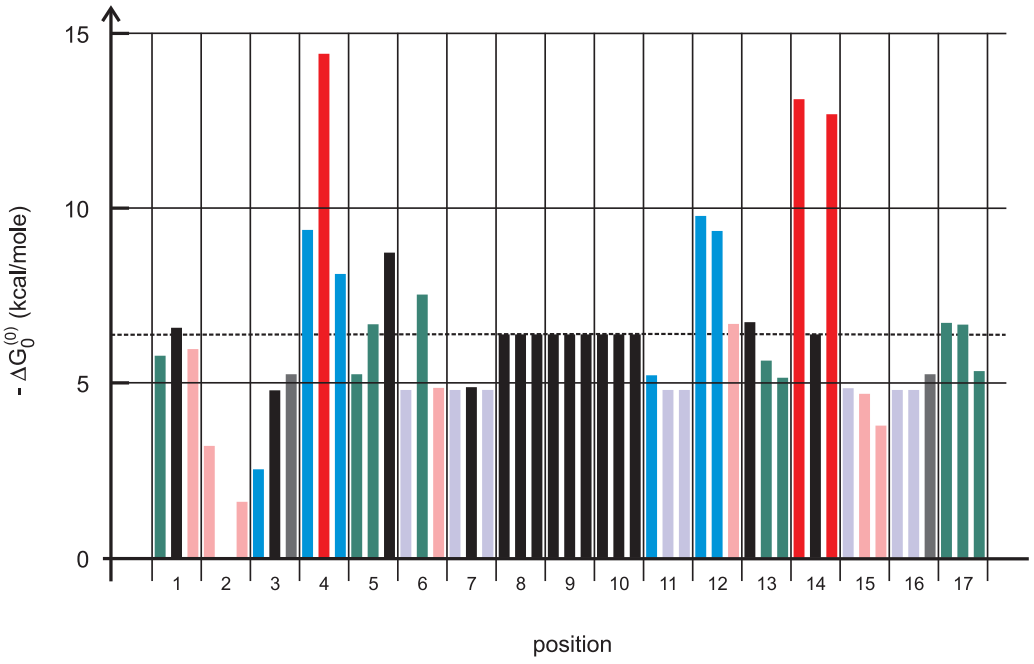


Figure 5.6: **Free energy landscape of a small RNA.** Free energies of folding, $-\Delta G_0^{(0)}$ at 0 °C, are plotted for the individual point mutations, which are grouped according to their positions along the sequence (from 5'- to 3'-end). The color code refers to the number of base pairs in the structures (see Fig.5.5): powder blue for two base pairs, pink for three base pairs, sky blue for four base pairs, grass green for five base pairs, black for six base pairs as in the reference structure, and red for seven base pairs. Mutations in the hairpin loop – positions 8, 9, 10 – do not change the structure. All calculations were performed with the Vienna RNA Package, version 1.8.5 [145].

stacks with three base pairs each (Fig.5.5; black structure), and the free energy of structure formation is: $\Delta G_0^{(0)} = -6.39\,\text{kcal/mole}$. The molecule is relatively stable and has three low lying suboptimal conformations with free energies at 0.29, 0.38, and 0.67 kcal/mole above the minimum free energy structure. These three states differ from the ground state through opening of one or both external base pairs (see e.g. Fig.5.7), all three suboptimal configurations are lying within reach of thermal energy and therefore contribute to the partition function of the molecule. The Hamming distance one neighborhood of $\mathbf{X}_0$ consists of $17 \times 3 = 51$ sequences, which form 16 different structures. Out of the 51 sequences 15 form the same minimum free energy structure as $\mathbf{X}_0$ and 10 have also the same minimum free energy energy implying a local degree of neutrality of $\lambda_0^{(0)} = 0.29$ for structures and $\lambda_0^{(\Delta G)} = 0.19$ for the free energies, respectively. The plot of the free energies of folding $\Delta G_0^{(0)}$ for all 51 mutants in Fig.5.6 is a perfect illustration of the ruggedness of the free energy landscape: The stability range of the one-error mutants goes from marginal stability at position 2, $\mathbf{G} \to \mathbf{U}$, to more than twice the absolute free energy of the reference at position 4, $\mathbf{U} \to \mathbf{G}$.

Recently, methods were developed that allow for efficient construction of fitness landscapes for catalytically active RNA molecules [235]. Ruggedness and neutrality are not restricted to RNA-molecules, similar results providing direct evidence were found with proteins [140]. Protein space, however, is more complex than RNA space since a large fraction of amino acid sequences does not lead to stable protein structures. Strongly hydrophobic molecular surfaces, for example, lead to protein aggregation and accordingly the landscapes have holes where no sequences are situated that cane give rise to useful structures. The landscape then reminds of a *holey landscape* as introduced in a different context by Sergey Gavrilets [109]. Holey landscapes provide a challenge for adaptive evolution because certain areas of sequence are not accessible. Attempts were made to reconstruct fitness landscapes for simple parasitic organisms like viruses. A recent example is large scale fitness modeling for HIV I [181]. Apart from a few exceptions experimental comprehensive information on fitness landscapes or conformational free energy surfaces is still rare but the amount of reliable data is rapidly grow-

ing. It seems to be appropriate therefore to conceive and construct model landscapes that account for the known features and to study evolutionary dynamics on them.

5.1.3 Sequences with multiple structures

The conventional view of biopolymer structures can be characterized as the one sequence-one structure paradigm: Given a sequence, the folding problem consists in the determination of the corresponding structure. A definition of the physical conditions under which the folding process occurs is required. Commonly, one assumes the thermodynamic minimum free energy criterium that is appropriate for small RNA molecules only. Large RNA molecules often exist and exert their function in long living metastable conformations, which were formed according to kinetic criteria, for example folding on-the-fly during RNA synthesis or kinetic folding of the entire sequence via kinetically determined folding nuclei. Here, we want to concentrate on RNA molecules that form multiple structures and function as RNA switches in nature and *in vitro*.

In the great majority of natural, evolutionary selected RNA-molecules we are dealing with sequences forming a single stable structure, whereas randomly chosen sequences generically form a great variety of metastable suboptimal structures in addition to the minimum free energy structure [248]. Important exceptions of the one sequence-one structure paradigm are RNA switches fulfilling regulatory functions in nature [122, 194, 260, 312] and synthetic biology [24]. Such riboswitches are multiconformational RNA molecules, which are involved in posttranscriptional regulation of gene expression. The conformational change is commonly induced by ligand binding or ribozymic RNA cleavage. Multitasking by RNA molecules clearly imposes additional constraints on genomic sequences and manifests itself through a higher degree of conservation in phylogeny.

The extension of the notion of structure to multiconformational molecules is sketched in Fig.5.7. The RNA molecule shown there has been designed to form two conformations: (i) a well defined minimum free energy (mfe)

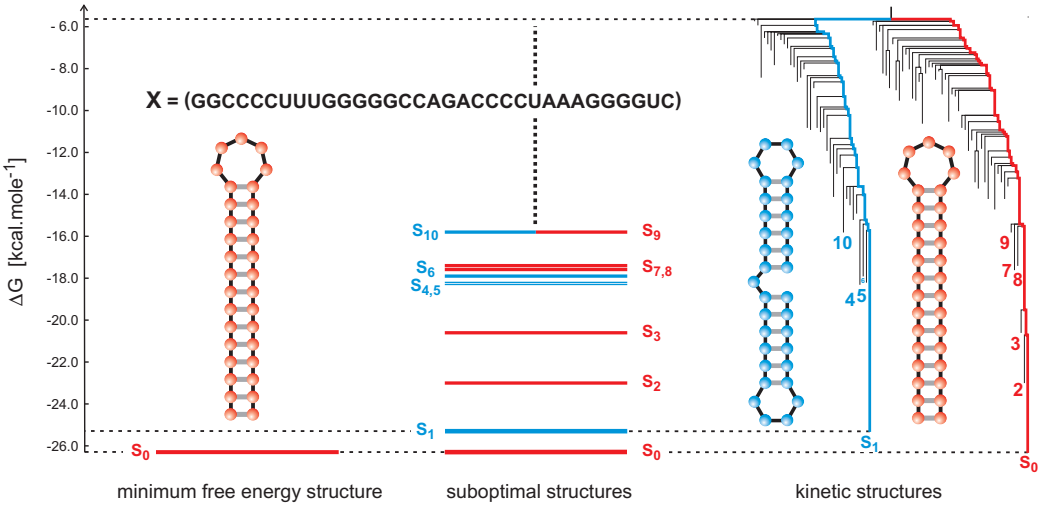


Figure 5.7: **RNA secondary structures viewed by thermodynamics and folding kinetics.** An RNA sequence $\mathbf{X}$ of chain length $l = 33$ nucleotides has been designed to form two structures: (i) the single hairpin mfe structure,$\mathbf{S}_0$ (red) and (ii) a double hairpin metastable structure, $\mathbf{S}_1$ (blue). The Gibbs free energy of folding (ΔG) is plotted on the ordinate axis. The leftmost diagram shows the minimum free energy structure $\mathbf{S}_0$ being a single long hairpin with a free energy of $\Delta G = -26.3$ kcal/mole. The plot in the middle contains, in addition, the spectrum of the ten lowest suboptimal conformations classified and color coded with respect to single hairpin shapes (red) and double hairpin shapes (blue). The most stable – nevertheless metastable – double hairpin has a folding free energy of $\Delta G = -25.3$ kacl/mole. The rightmost diagram shows the barrier tree of all conformations up to a free energy of $\Delta G = -5.6$ kcal/mole where the energetic valleys for the two structures merge into one basin containing 84 structures, 48 of them belonging to the single hairpin subbasin and 36 to the double hairpin subbasin. A large number of suboptimal structures has free energies between the merging energy of the subbasins and the free reference energy of the open chain ($\Delta G = 0$).

structure – being a perfect single hairpin (red structure; lhs of the figure and rhs of the barrier tree), and (ii) a metastable double hairpin conformation (blue structure; lhs of the barrier tree), which is almost as stable as the mfe structures. In addition to the mfe structure $\mathbf{S}_0$ and the conformation $\mathbf{S}_1$, the sequence $\mathbf{X} \equiv (\mathbf{GGCCCCUUUGGGGGCCAGACCCCUAAAGGGGUC})$ like almost all RNA sequences[6] forms a great variety of other, less stable

[6]Exceptions are only very special sequences, homopolynucleotides, for example.

conformations called *suboptimal structures* ($\mathbf{S}_2$ to $\mathbf{S}_{10}$ are shown by numbers in the middle of the figure and in the barrier tree). The numbers of all suboptimal structures are huge but most suboptimal conformations have free energies way above the mfe structure and don't appreciably contribute to the low-lying conformations. Structures, mfe and suboptimal structures, are related through transitions, directly or via intermediates, which in a simplified version can be represented by means of a barrier tree [95, 313] shown on the rhs of the figure. Kinetic folding introduces a second time scale into the scenario of molecular evolution.[7] Based on Arrhenius theory of chemical reaction rates,

$$k \;=\; A \cdot e^{-E_{\rm a}/RT} \;, \tag{5.5}$$

the height of the barrier, $E_{\rm a}$ determines the reaction rate parameter k and thereby the half life of the conformation $t_{1/2} = \ln 2/k$. In equation (5.5), A is the pre-exponential factor of the reaction, R is the gas constant and T the absolute temperature in $^\circ$Kelvin. The two structures shown in Fig.5.7 are connected by a lowest barrier of 20.7 kcal/mole that depending on the pre-exponential factors implies half lives of days or even weeks for the two conformations. In a conventional experiment with a time scale of hours the two conformations would appear as two separate entities. Barriers, nevertheless, can be engineered to be much lower and then an equilibrium mixture of rapidly interconverting conformations may be observed. Several constraints are required for the conservation of an RNA switch, and the restrictions of variability in sequence space are substantial.

The comparison of the two dominant structures $\mathbf{S}_0$ and $\mathbf{S}_1$ in Fig. 5.7 provides a straightforward example for the illustration of different notions of stability: (i) thermodynamic stability, which considers only the free energies of the mfe structures – $\mathbf{S}_0$ in Fig.5.7 is more stable than $\mathbf{S}_1$ since it has a lower free energy, $\Delta G(\mathbf{S}_0) < \Delta G(\mathbf{S}_1)$, (ii) conformational stability, which can be expressed in terms of suboptimal structures or partition functions within a

[7] Timescale number one is the evolutionary process – mutation and selection – itself. In order to be relevant for evolutionary dynamics the second timescale has to be substantially faster than the first one, or in other words, the conformational changes has to occur almost instantaneously on the timescale of evolution.

basin or a subbasin of an RNA sequence – a conformationally stable molecule has no low lying suboptimal conformations that can be interconverted with the mfe structure or another metastable structure at the temperature of the experiment; for kinetic structures separated by high barriers the partition functions are properly restricted to individual subbasins [195], and (iii) mutational stability that is measured in terms of the probability with which a mutation changes the structure of a molecule (see Fig.5.2). All three forms of stability are relevant for evolution, but mutational stability and the spectrum of mutational effects – adaptive, neutral or deleterious – are most important.

RNA suboptimal structures were also used for explaining characteristic feature of the evolution of organisms like plasticity, evolvability, and modularity [6, 7, 97]. Among these very complex features modularity is the most intriguing one, because it plays a crucial role from the early beginnings of evolution to the most complex relations in societies. Modularity is indispensable for understanding complex systems in biology and anywhere else. Modular structure in cases where it is not congruent with other forces shaping subunits is hard to interpret still and as it seems will remain a hot topic for many years in the future.

5.2 Dynamics on realistic rugged landscapes

The majority of data on the relation between sequences and molecular properties comes from structural biology of biopolymers, in particular RNA and protein. RNA secondary structures as shown in the previous section 5.1 provide a simple and mathematically accessible example of a *realistic* mapping of biopolymer sequences onto structures [252]. The RNA model is commonly restricted to the assignment of a single structure to every sequence but the explicit consideration of suboptimal conformations is possible as well (see subsection 5.1.3 and [248]). Two features are characteristic for landscapes derived from RNA molecules: (i) *ruggedness* – pairs of sequences situated nearby in sequence space, i.e., having Hamming distance $d_{\mathrm{H}} = 1$, may give rise to very similar or entirely different structures and properties – and (ii)

neutrality – two or more sequences may have identical structures and properties.

Rugged fitness landscapes, which are more elaborate than the simple ones discussed in section 4.3.6, have been proposed. The most popular example is the Nk-model conceived by Stuart Kauffman [169, 170, 306] that is based on individual loci on a genome and interactions between them: N is the number of loci and k is the number of interactions. A random element, which is drawn from a predefined probability distribution – commonly the normal distribution – and which defines the interaction network, is added to the otherwise deterministic model: N and k are fixed and not subjected to variation. Here a different approach is proposed that starts out from the nucleotide sequence of a genome rather than from genes and alleles, and consequently it is based on the notion of sequence space. Ruggedness (this section 5.2) and neutrality (see section 5.3) are introduced by means of tunable parameters, d and λ, and pseudorandom numbers are used to introduce random scatter, which reflects the current ignorance with respect to detailed fitness values and which is thought to be replaced by real data when they become available in the near future. We begin with an overview of the current knowledge on biopolymer landscapes and discuss afterwards model landscapes that come close to real landscapes at the current state of knowledge and investigate then evolutionary dynamics on such "realistic" landscapes.

A new type of landscapes, the *realistic rugged landscape* (RRL), is introduced and analyzed here. Ruggedness is modeled by assigning fitness differences at random within a predefined band of fitness values with adjustable width d. The highest fitness value is attributed to the master sequence $\mathbf{X}_m \doteq \mathbf{X}_0$, $f_m = f_0$, and the fitness values of all other sequences are obtained by means of the equation

$$f(\mathbf{X}_j) \;=\; f_j \;=\; \begin{cases} f_0 & \text{if } j = 0 \;, \\ f + 2d(f_0 - f)\left(\eta_j^{(s)} - 0.5\right) & \text{if } j = 1, \ldots, \kappa^l \;, \end{cases} \tag{5.6}$$

where $\eta_j^{(s)}$ is the j-th output random number from a pseudorandom number

generator with a uniform distribution of numbers in the range $0 \leq \eta_j^{(s)} \leq 1$. The random number generator is assumed to have been started with the seed s,[8] which will be used to characterize a particular distribution of fitness values (Fig. 5.8). The parameter d determines the amount of scatter around the mean value $\bar{f}_{-0} = f$, which is independent of d: $d = 0$ yields the single peak landscape, and $d = 1$ leads to fully developed or maximal scatter where individual fitness values f_j can reach the value f_0. A given landscape can be characterized by

$$\mathcal{L} \;=\; \mathcal{L}(\lambda, d, s; l, f_0, f)\;, \tag{5.7}$$

where λ is the degree of neutrality (see section 5.3; here we have $\lambda = 0$). The parameters l, f_0 and f have the same meaning as for the single peak landscape (3.4c).

Two properties of realistic rugged landscapes fulfilled by fitness values relative to the mean except the master, $\varphi_j = f_j - f \,\forall\, j = 0, \ldots, \kappa^l - 1$, are important: (i) the ratio of two relative fitness values of sequences within the mutant cloud is independent of the scatter d and (ii) the ratio of the relative fitness values of a sequence from the cloud and the master sequence is proportional to the scatter d:

$$\frac{\varphi_j}{\varphi_k} \;=\; \frac{\eta_j^{(s)} - 0.5}{\eta_k^{(s)} - 0.5}\,;\; j,k = 1, \ldots, \kappa^l - 1 \;\text{ and} \tag{5.8a}$$

$$\frac{\varphi_j}{\varphi_0} \;=\; 2\,d\,\left(\eta_j^{(s)} - 0.5\right)\,;\; j = 1, \ldots, \kappa^l - 1\;. \tag{5.8b}$$

The second equation immediately shows that $\sum_{j=1}^{\kappa^l - 1} \varphi_j = 0$.

5.2.1 Single master quasispecies

We are now in a position to explore whether or not the results derived from simple model landscapes are representative for mutation-selection dynamics in real populations. At first the influence of random scatter on quasispecies

[8] The seed s indeed determines all details of the landscape, which is completely defined by s and the particular type of the pseudorandom number generator as well as by f_0, f, and d.

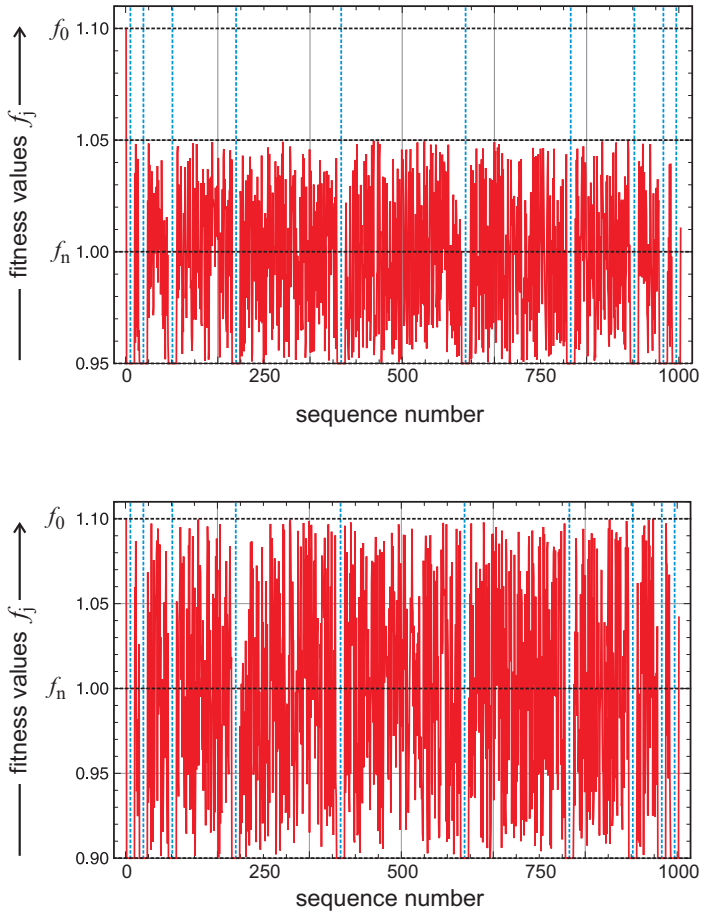


Figure 5.8: ***Realistic* rugged fitness landscapes.** The landscapes for binary sequences with chain length $l = 10$ are constructed according to equation (5.6). In the upper plot the band width of random scatter was chosen to be $d = 0.5$ and a seed of $s = 919$ was used for the random number generator ($\mathcal{L}(0, 0.5, 919; 10, 1.1, 1.0)$). For the lower plot showing maximal random scatter $d = 1$ and $s = 637$ was applied ($\mathcal{L}(0, 1.0, 637; 10, 1.1, 1.0)$). Careful inspection allows for the detection of individual differences. The broken blue lines separate different mutant classes.

and error thresholds will be studied. The chain length for which diagonalization of the value matrix W can be routinely performed lies at rather small values around $l = 10$ giving rise to a matrix size of $1\,000 \times 1\,000$. Accord-

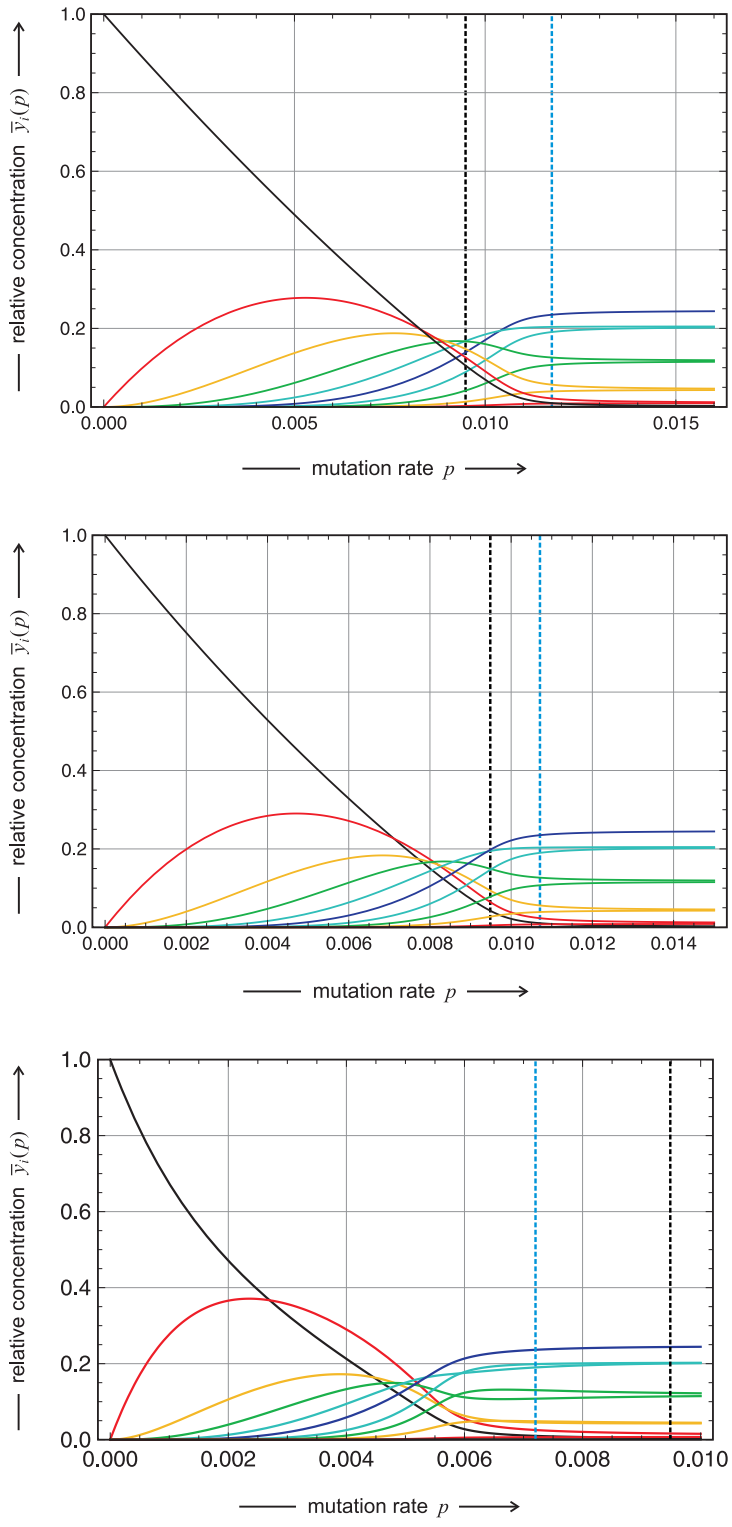


Figure 5.9: **Error thresholds on a *realistic* model landscape with different random scatter d.** Shown are the stationary concentrations of classes $\bar{y}_j(p)$ on the *realistic* landscape with $s = 023$ for $d = 0$ ($\mathcal{L}(0, 0, 0)$; upper plot), $d = 0.5$ ($\mathcal{L}(0, 0.5, 023)$; middle plot), and $d = 0.95$ ($\mathcal{L}(0, 0.95, 023)$; lower plot). The error threshold calculated by zero mutational backflow lies at $p_{\text{cr}} = 0.009486$ (black dotted line), the values for level crossing decrease with the width of random scatter d (blue dotted lines). Other parameters: $l = 10$, $f_0 = 1.1$, and $f = 1.0$.

ingly, it has to be confirmed first whether or not such a short chain length is sufficient to yield representative results. In Fig. 5.9 the stationary concentrations of mutant classes, $\bar{y}_k$ $(k = 0, 1, \ldots, 10)$ are shown for different band widths d of random scatter: The purely deterministic case $d = 0$ representing the single-peak landscape, $d = 0.5$, and $d = 0.95$, the maximal scatter that sustains a single quasispecies over the entire range, $0 \leq p < p_{\mathrm{cr}}$.[9] Despite the short chain length of $l = 10$ the plots reflect the threshold phenomena rather well, the width of the transition to the uniform distribution is hardly changing, and values for level crossing (section 4.3.6 and table 4.3) are shifted towards smaller $p_{(1/M)}$-values with increasing d. Answering the initial question, computer studies with $l = 10$ are suitable for investigations on quasispecies behavior.

For $d > 0$ the fitness values for individual sequences within one class are no longer the same and hence the curves $\bar{x}_j(p)$ differ from each other and form a band for each class that increases in width with the amplitude d of the random component (Fig. 5.11). The separation of the bands formed by curves belonging to different error classes is always recognizable at sufficiently small mutation rates p but the bands overlap and merge at higher p-values. As expected the zone where the bands begin to mix moves in the direction $p = 0$ with increasing scatter d. Interestingly, the error threshold phenomenon is fully retained thereby, only the level-crossing value $p_{(1/100)}$ is shifted towards lower error rates (figs. 5.10, 5.11, and 5.12). Indeed, the approaches towards the uniform distribution on the landscape without a random component $(d = 0)$ and on the landscape with $d = 0.5$ are very similar apart from the relatively small shift towards lower p-values, whereas the shift for $d = 0.95$ is substantially larger and the solution curve $\bar{x}_0(p)$ is curved upwards more strongly. Closer inspection of the shift of the level-crossing value shows nonmonotonous behavior for some landscapes: The level crossing value is shifted towards larger p-values at first, passes a maximum value and then

[9] As shown below in detail (Fig. 5.14) individual quasispecies may be replaced by others at certain critical p-values, p_{tr}. For a given scatter s the number of such transitions becomes larger with increasing values of d.

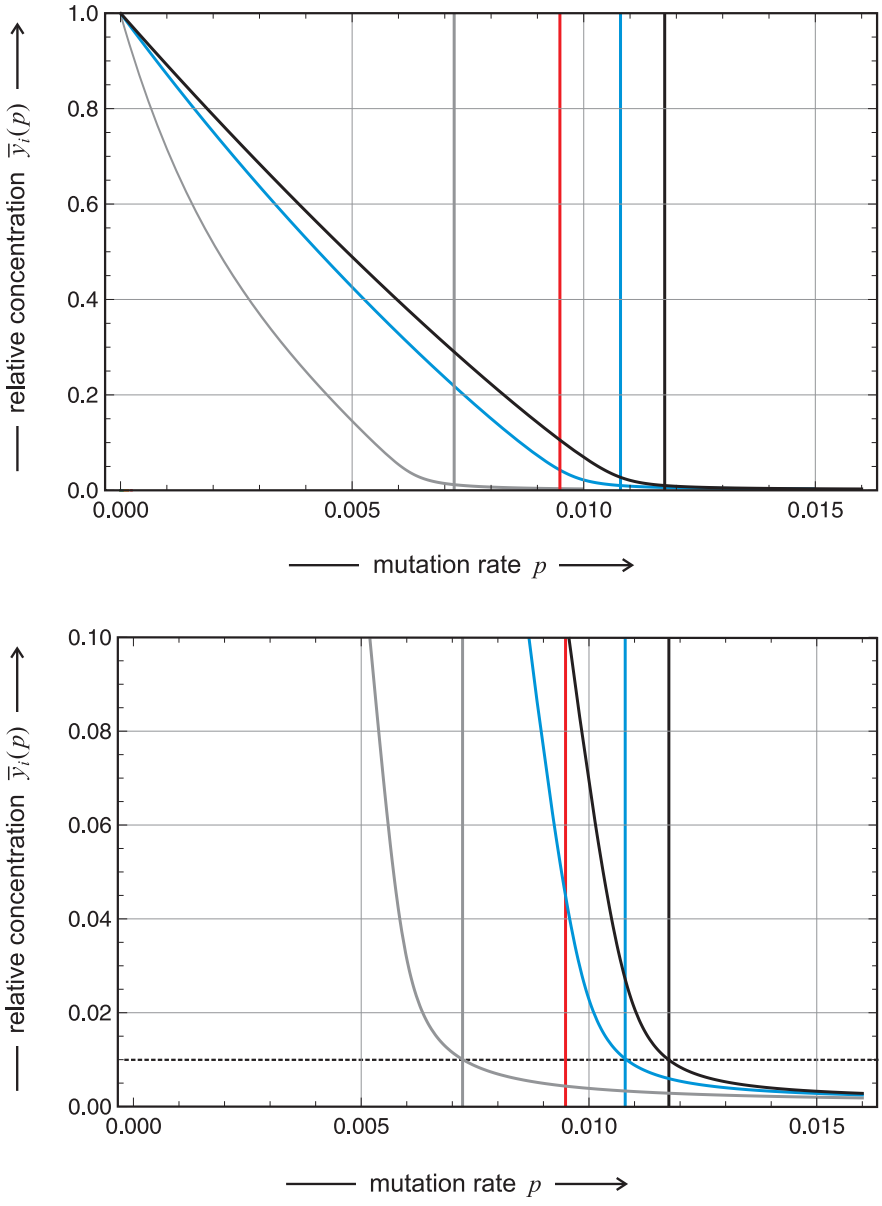


Figure 5.10: **Error threshold and decay of the master sequence $\mathbf{X}_0$.** Shown are the stationary concentrations of the master sequence $\bar{x}_0(p)$ and the level crossing values $p_{(1/100)}$ (vertical lines) on a landscape with $s = 023$ for $d = 0$ (black), $d = 0.5$ (blue), and $d = 0.950$ (grey). The error threshold lies at $p_{\mathrm{cr}} = 0.094857$ (red). The lower plot enlarges the upper plot and shows the level $\bar{x}_0 = 0.01$ (dotted horizontal line, black). Other parameters: $l = 10$, $f_0 = 1.1$ and $f = 1.0$.

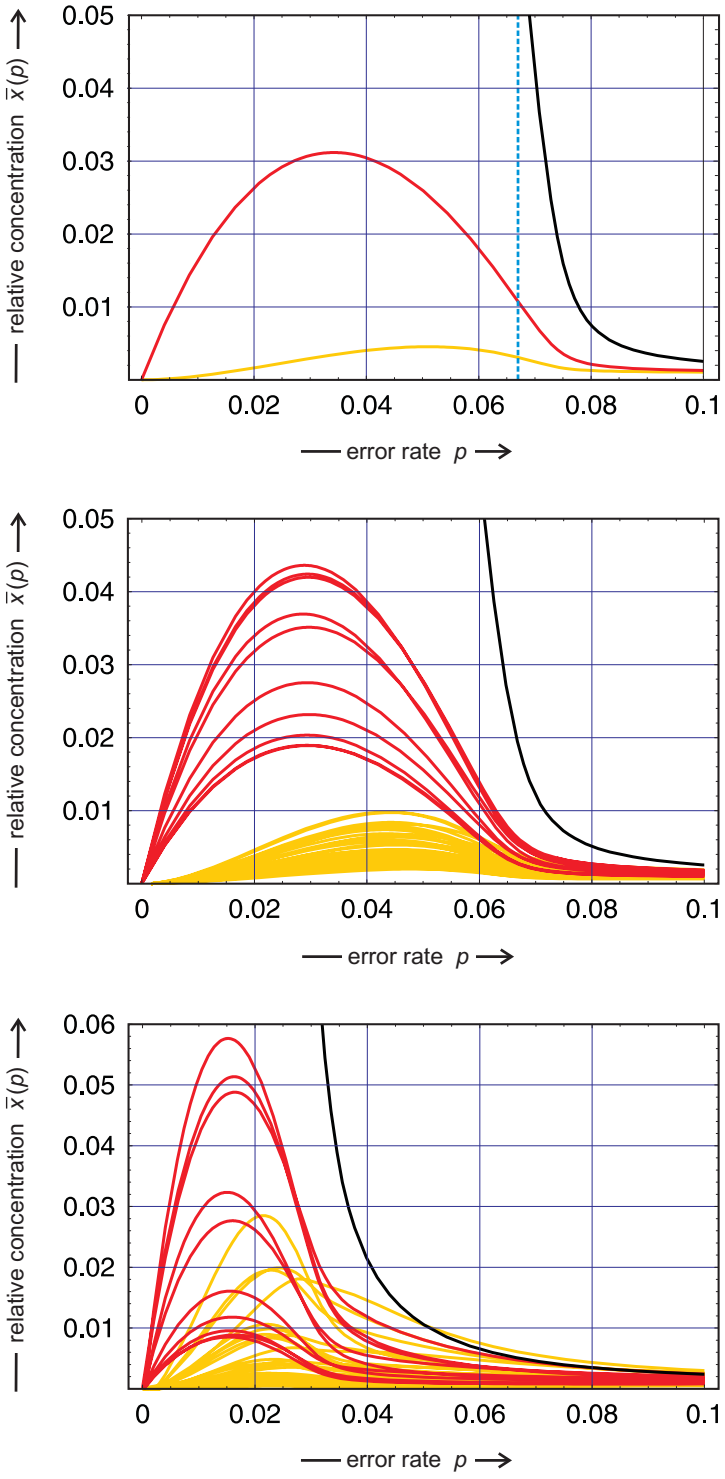


Figure 5.11: **Quasispecies on a *realistic* model landscape with different random scatter** d**.** Shown are the stationary concentrations $\bar{x}_j(p)$ on a landscape with $s = 491$ for $d = 0$ (upper plot), $d = 0.5$ (middle plot), and $d = 0.9375$ (lower plot) for the classes Γ_0, Γ_1, and Γ_2. In the topmost plot the curves for all sequences in Γ_1 (single point mutations, $d_{\mathrm{H}}(\mathbf{X}_0, \mathbf{X}_{(1)}) = 1$) coincide, and so do the curves in Γ_2 (double point mutations, $d_{\mathrm{H}}(\mathbf{X}_0, \mathbf{X}_{(2)}) = 2$) since zero scatter, $d = 0$, has been chosen. The error threshold calculated by zero mutational backflow lies at

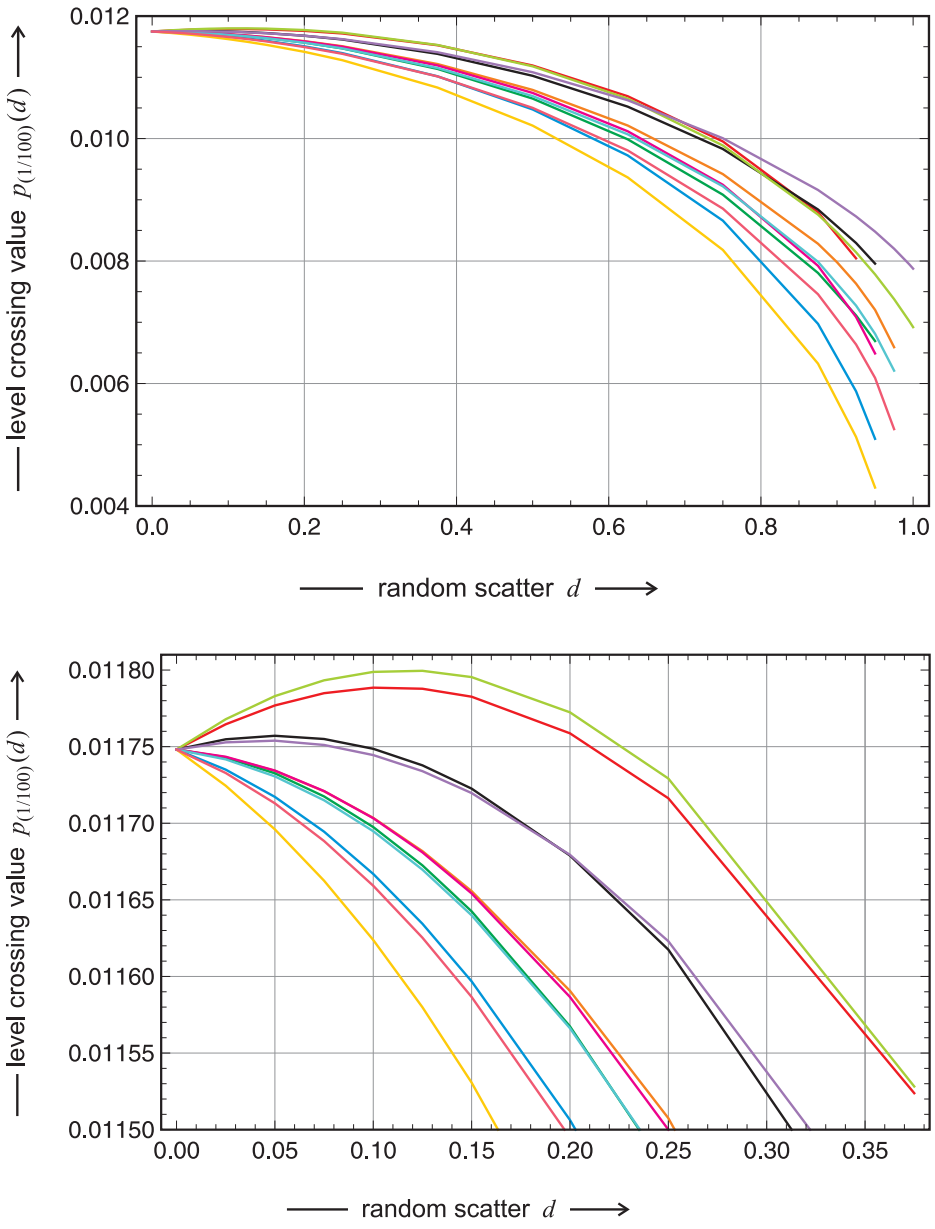


Figure 5.12: **Level-crossing values for the master sequence of different model landscape with different random scatter** d**.** Shown are the level crossing values for $M = 100$ as functions of the random scatter $p_{(1/100)}(d)$. The error threshold calculated by zero mutational backflow lies at $p_{\mathrm{cr}} = 0.0117481$. Color code for different seeds s: 023 = orange, 229 = red, 367 = green, 491 = black, 577 = chartreuse, 637 = blue, 673 = yellow, 877 = magenta, 887 = turquoise, 919 = blue violet, and 953 = hot pink. Other parameters: $l = 10$, $f_0 = 1.1$, and $f = 1.0$.

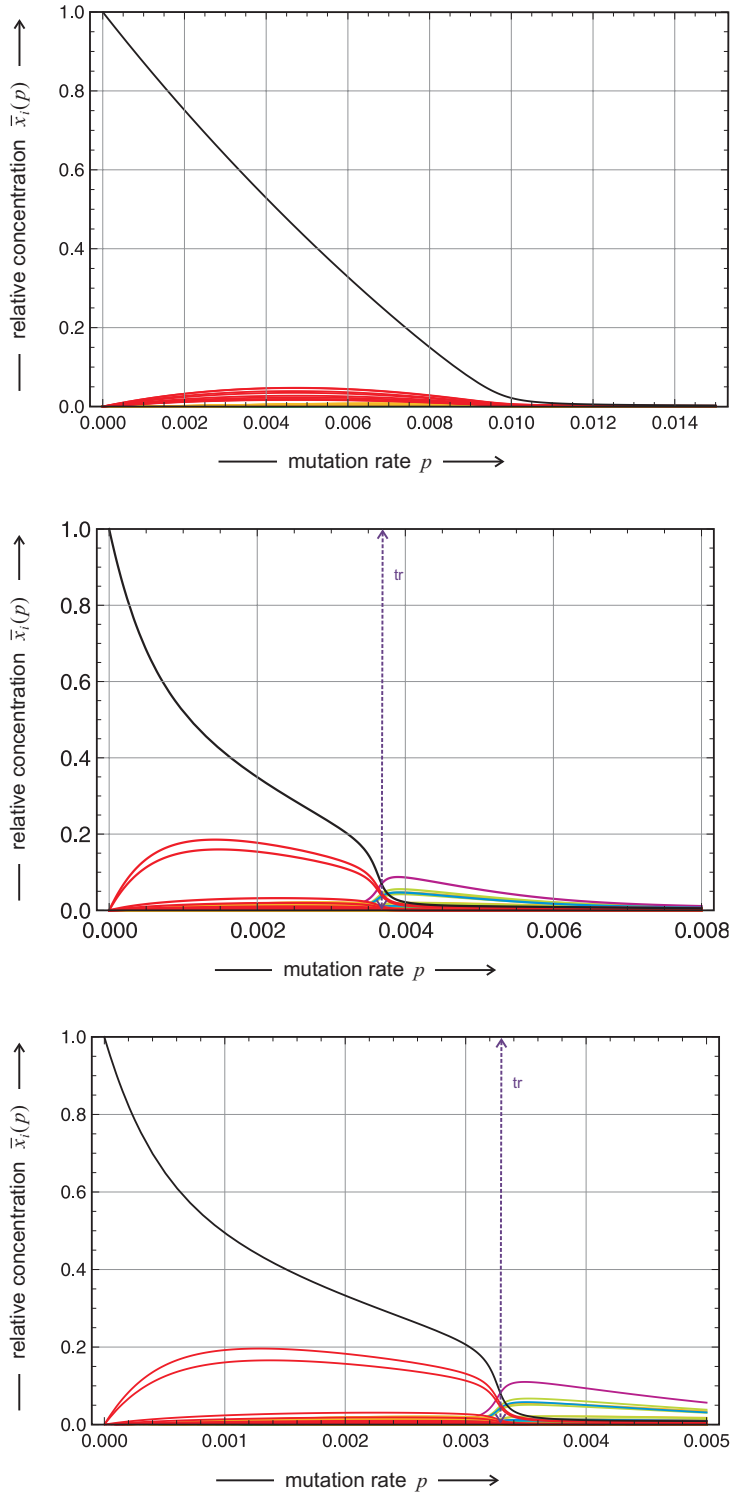


Figure 5.13: **A *realistic* model landscape with a transition between quasispecies.** Shown are the stationary concentrations $\bar{x}_j(p)$ on a landscape with $s = 023$ for $d = 0.5$ (upper plot), $d = 0.999$ (middle plot), and fully developed scatter $d = 1.0$ (lower plot). Other parameters: $l = 10$, $f_0 = 1.1$, and $f = 1.0$.

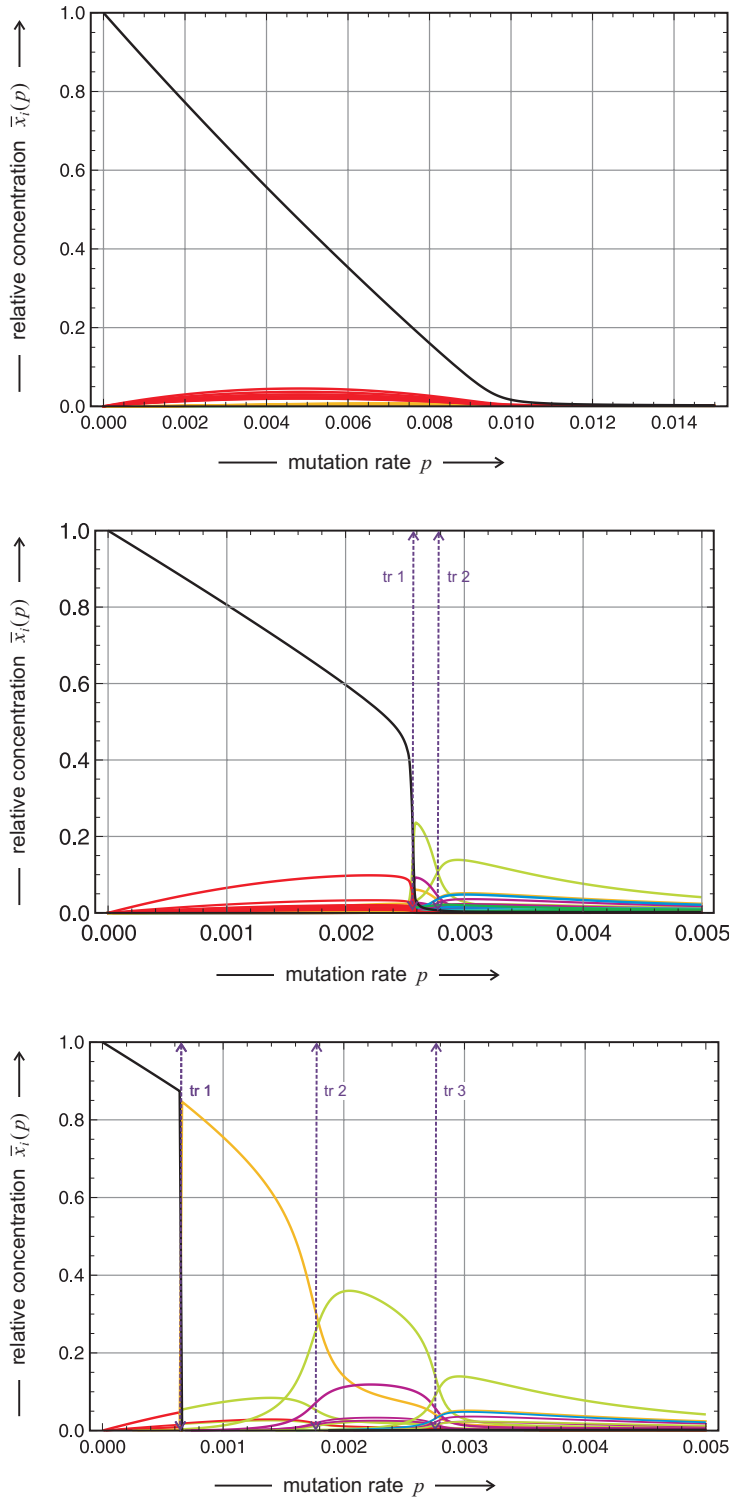


Figure 5.14: **A *realistic* model landscape with multiple transitions between quasispecies.** Shown are the stationary concentrations $\bar{x}_j(p)$ on a landscape with $s = 637$ for $d = 0.5$ (upper plot), $d = 0.995$ (middle plot), and fully developed scatter $d = 1.0$ (lower plot). Other parameters: $l = 10$, $f_0 = 1.1$, and $f = 1.0$.

follows the general shift towards lower values of p with increasing scatter d (Fig. 5.12).

5.2.2 Transitions between quasispecies

In addition, transitions between quasispecies may be observed at critical mutation rates $p = p_{\text{tr}}$: One quasispecies, $\bar{\Upsilon}_0$, which is the stationary solution of the mutation-selection equation (4.9) in the range $0 \leq p < p_{\text{tr}}$, is replaced by another quasispecies, $\bar{\Upsilon}_k$, which represents the stationary solution above the critical value up to the error threshold $p_{\text{tr}} < p < p_{\text{cr}}$, or up to a second transition, $(p_{\text{tr}})_1 < p < (p_{\text{tr}})_2$. More than two transitions are also possible, an example is shown in Fig. 5.14 (lower plot). The mechanism by which quasispecies replace each other is easily interpreted [256]:[10] The stationary mutational backflow from the sequences $\mathbf{X}_i$ $(i = 1, \ldots, n)$ to the master sequence $\mathbf{X}_0$ is determined by the sum of the product terms $\psi_0 = \sum_{i=1}^{n} W_{0i} = Q_{0i} f_i$ and likewise for a potential master sequence $\mathbf{X}_k$, $\psi_k = \sum_{i=0, i\neq k}^{n} W_{ki} = Q_{ki} f_i$. The necessary – but not sufficient – condition for the existence of a transition is $\Delta\psi = \psi_0 - \psi_k < 0$. Since the fitness value f_0 is the largest by definition we have $f_0 > f_i$ $(i = 1, \ldots, n)$ and at sufficiently small mutation rates p the differences in the values, $\Delta\omega = \omega_0 - \omega_k = W_{00} - W_{kk} = Q_{00} f_0 - Q_{kk} f_k > 0$, will always outweigh the difference in the backflow, $\Delta\omega > |\Delta\psi|$. With increasing values of p, however, the replication accuracy and $\Delta\omega$ will decrease because of the term $Q_{00} = Q_{kk} \approx (1-p)^l$ in the uniform error approximation. At the same time $\Delta\psi$ will increase in absolute value and provided $\Delta\psi < 0$ there might exist a mutation rate $p = p_{\text{tr}}$ smaller than the threshold value $p_{\text{tr}} < p_{\text{cr}}$ at which the condition $\Delta\omega + \Delta\psi = 0$ is fulfilled and consequently, the quasispecies $\bar{\Upsilon}_k$ is the stable stationary solution of equation (4.9) at $p > p_{\text{tr}}$. The influence of a distribution of fitness values instead of the single value of the single-peak landscapes can be predicted straightforwardly: Since f_0 is independent of the fitness scatter d the difference $f_0 - f_k$ will decrease with

[10] Thirteen years after this publication the phenomenon has been observed in quasispecies of digital organisms [309] and was called *survival of the flattest*.

Table 5.1: **Computed and numerically observed quasispecies transitions.** In the table we compare the numerically observed values of p at transitions between quasispecies, p_{tr}, with the values calculated from equation (5.9), $(p_{\mathrm{tr}})_{\mathrm{eval}}$, and the error thresholds, p_{cr}. The table is adopted from [256].

Chain length	Qsp. $\bar{\Upsilon}_0$		Qsp. $\bar{\Upsilon}_m$		Critical mutation rates		
l	f_0	$f_1^{(0)}$	f_m	$f_1^{(m)}$	p_{tr}	$(p_{\mathrm{tr}})_{\mathrm{eval}}$	p_{cr}
20	10	1	9.9	2	0.0520	0.0567	0.1130
50	10	1	9.9	2	0.0362	0.0366	0.0454
50	10	1	9.9	5	0.0148	0.0147	0.0470
50	10	1	9.0	5	0.0445	0.0456	0.0453

increasing scatter d. Accordingly, the condition for a transition between quasispecies can be fulfilled at lower p-values and we expect to find one or more transitions preceding the error threshold p_{cr}. No transition can occur on the single peak landscape but as d increases and the difference $\Delta\omega$ becomes smaller and it becomes more likely that the difference in backflow becomes sufficiently strong for a replacement of $\bar{\Upsilon}_0$ by $\bar{\Upsilon}_k$ below p_{cr}. Fig. 5.13 presents a typical example: No quasispecies transition is found up to a random scatter of $d = 0.95$. Then, a soft transition becomes observable at $d = 0.975$ and eventually dominates the plot of the quasispecies against the mutation rate p at random scatter close to the maximum ($d = 0.995$ and $d = 1.0$). An example with multiple transitions increasing in number with increasing d is shown in Fig. 5.14.

An explicit computation of the transition point $p = p_{\mathrm{tr}}$ has been performed some time ago [256]. A simple model is used for the calculation of the critical value that is based on a zero mutational flow assumption between the two quasispecies. The value matrix W corresponding to all 2^l sequences of chain length l is partitioned into two diagonal blocks and the rest of the matrix: (i) Block $\bar{\Upsilon}_0$ contains sequence $\mathbf{X}_0$ with the highest fitness value f_0, which is the master sequence in the range $0 \leq p < p_{\mathrm{tr}}$ and all its one-error

mutants $\mathbf{X}_{(1)}^{(0)} = \{\mathbf{X}_j \in \Gamma_1^{(0)}\}$ with a fitness value $f_1^{(0)}$, (ii) block $\bar{\Upsilon}_m$ contains sequence $\mathbf{X}_m$ with the fitness value f_m, which is the master sequence in the range $p_{\text{tr}} \leq p < p_{\text{cr}}$, and all its one-error mutants $\mathbf{X}_{(1)}^{(m)} = \{\mathbf{X}_j \in \Gamma_1^{(m)}\}$ with a fitness value $f_1^{(m)}$, and (iii) the rest of the matrix W is neglected completely as all entries are set equal zero:

$$W = q^l \begin{pmatrix} f_0 & f_1^{(0)}\varepsilon & \cdots & f_1^{(0)}\varepsilon & \cdots & 0 & 0 & \cdots & 0 \\ f_0\varepsilon & f_1^{(0)} & \cdots & f_1^{(0)}\varepsilon^2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_0\varepsilon & f_1^{(0)}\varepsilon^2 & \cdots & f_1^{(0)} & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & f_m & f_1^{(m)}\varepsilon & \cdots & f_1^{(m)}\varepsilon \\ 0 & 0 & \cdots & 0 & \cdots & f_m\varepsilon & f_1^{(m)} & \cdots & f_1^{(m)}\varepsilon^2 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & f_m\varepsilon & f_1^{(m)}\varepsilon^2 & \cdots & f_1^{(m)} \end{pmatrix} .$$

Each block is now represented by a 2×2 matrix

$$W_0 = q^l \begin{pmatrix} f_0 & l\, f_1^{(0)}\varepsilon \\ f_0\varepsilon & f_1^{(0)}\big(1 + (l-1)\varepsilon^2\big) \end{pmatrix} \quad \text{and}$$

$$W_m = q^l \begin{pmatrix} f_m & l\, f_1^{(m)}\varepsilon \\ f_m\varepsilon & f_1^{(m)}\big(1 + (l-1)\varepsilon^2\big) \end{pmatrix} .$$

Calculation of the two largest eigenvalues λ_0 and λ_m yields the condition for the occurrence of the transitions: $\lambda_0 = \lambda_m$. The result is

$$p_{\text{tr}} = 1 - \sqrt{1 - \frac{\vartheta_{\text{tr}}}{l-1}} , \tag{5.9}$$

with ϑ_{tr} being the result of the equation

$$\vartheta_{\text{tr}} = \frac{1}{2}\Big(\alpha + \beta - \gamma + \sqrt{(\alpha + \beta - \gamma)^2 - 4\alpha\beta}\Big) \quad \text{with} \tag{5.9'}$$

$$\alpha = l - \frac{f_0 - f_m}{f_1^{(m)} - f_1^{(0)}} ,$$

$$\beta = l - \frac{f_0 f_m (f_1^{(m)} - f_1^{(0)})}{f_1^{(0)} f_1^{(m)} (f_0 - f_m)} , \quad \text{and}$$

$$\gamma = \frac{(f_0 f_1^{(0)} - f_m f_1^{(m)})^2}{(l-1) f_1^{(0)} f_1^{(m)} (f_0 - f_m)(f_1^{(m)} - f_1^{(0)})} .$$

Although the complexity of these equations is prohibitive for further manipulations, the accuracy of the zero mutational flow approximations is relevant for the next subsection 5.2.3 were we shall apply a similar approximation. The corresponding table 5.1 is reproduced from [256]. The agreement is very good indeed but is cases where p_{tr} and p_{cr} are very close it can nevertheless happen that the calculated value lies above the error threshold.

5.2.3 Clusters of coupled sequences

A certain fraction of landscapes gives rise to characteristic quasispecies distributions as a function of the mutation rate p that is substantially different from the one shown in Fig. 5.14 and discussed above. No transitions are observed, not even at fully developed scatter $d = 1$ (Fig. 5.15). Another feature concerns the classes to which the most frequent sequences belong. On the landscape defined by $s = 919$ these sequences are the master sequence ($\mathbf{X}_0$; black curve), one one-error mutant ($\mathbf{X}_4$; red curve), and one two-error mutant ($\mathbf{X}_{516}$; yellow curve).[11] The three sequences are situated close-by in sequence space – Hamming distances $d_{\mathrm{H}}(\mathbf{X}_1, \mathbf{X}_4) = d_{\mathrm{H}}(\mathbf{X}_4, \mathbf{X}_{516}) = 1$ and $d_{\mathrm{H}}(\mathbf{X}_1, \mathbf{X}_{516}) = 2$)– form a cluster, which is dynamically coupled by means of strong mutational flow (Fig. 5.16). Apparently, such a quasispecies is not likely to be replaced in a transition by another one that is centered around a single master sequence and accordingly, we call such clusters *strong quasispecies*. The problem that ought to be solved now is the prediction of the occurrence of strong quasispecies from know fitness values.

First, a heuristic is mentioned that serves as an (almost perfect) diagnostic tool for detecting whether or not a given fitness landscape gives rise to a strong quasispecies: (i) For every mutant class we identify the sequence with the highest fitness value, f_0, $(f_{(1)})_{\max} = f(\mathbf{X}_{m(1)})$, $(f_{(2)})_{\max} = f(\mathbf{X}_{m(2)})$, $\ldots$, and call them *class-fittest* sequences. Next we determine the fittest sequences in the one-error neighborhood of the class-fittest sequences. Clearly, for the class k-fittest sequence $\mathbf{X}_{m(k)}$ this sequence lies either in class $k-1$ or in class $k+1$.[12] Simple combinatorics is favoring classes closer to the middle of

[11] Naïvely we would expect a band of one-error sequences at higher concentration than the two-error sequence.

[12] For class $k = 1$ we omit the master sequence $\mathbf{X}_0$, which trivially is the fittest sequence

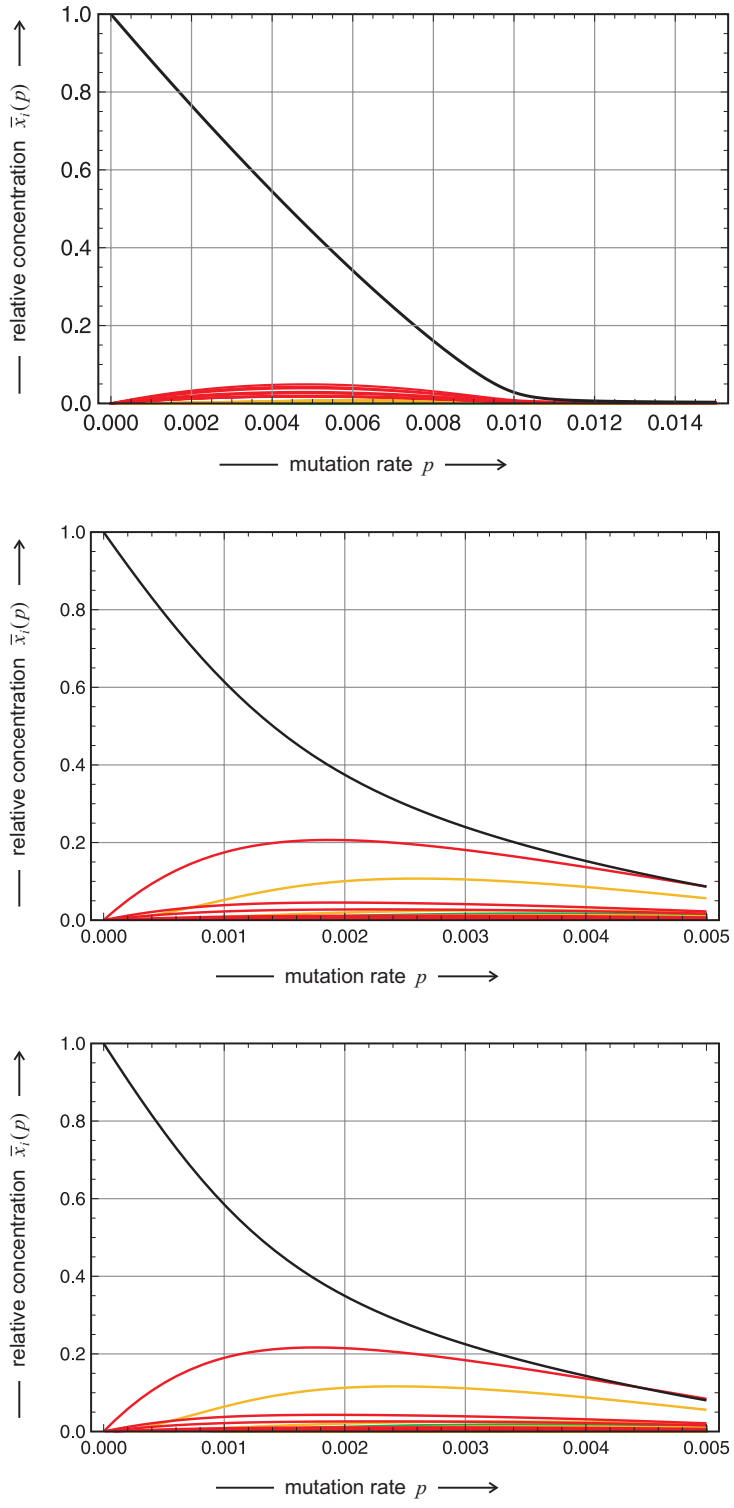


Figure 5.15: **Error thresholds on a *realistic* model landscape with different random scatter d and transitions between quasispecies.** The landscape characteristic is $s = 919$. Shown are the stationary concentrations $\bar{x}_j(p)$ for $d = 0.5$ (upper plot), $d = 0.995$ (middle plot), and fully developed scatter $d = 1.0$ (lower plot). Other parameters: $l = 10$, $f_0 = 1.1$, and $f = 1.0$.

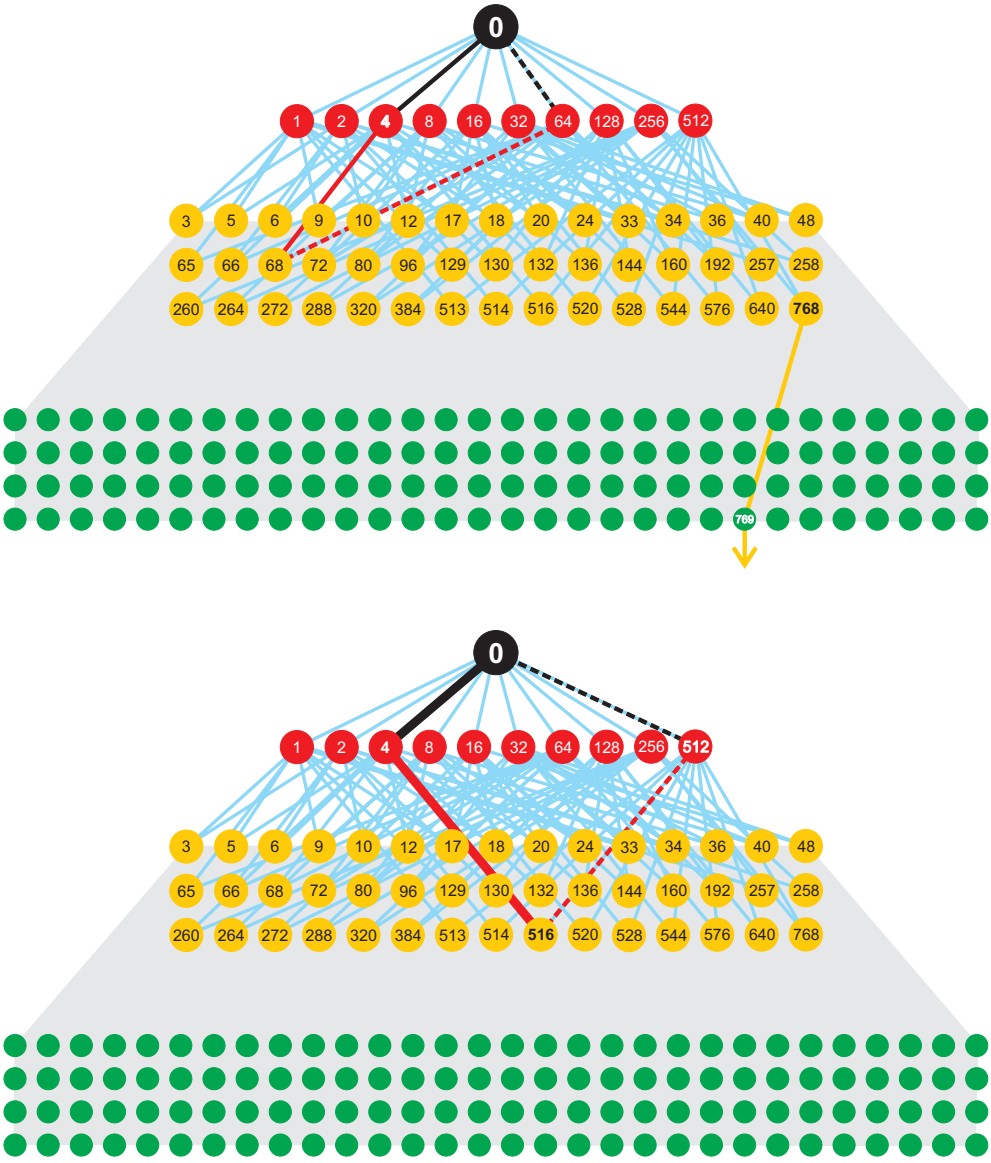


Figure 5.16: **Mutation flow in quasispecies.** The sketch shows two typical situations in the distribution of fitness values in sequence space. In the upper diagram ($s = 637$) the fittest two-error mutant, $\mathbf{X}_{768}$, has its fittest nearest neighbor, $\mathbf{X}_{769}$, in the three-error class Γ_3, and the fittest sequence in the one-error neighborhood of $\mathbf{X}_4$ (being the fittest sequence in the one-error class), $\mathbf{X}_{68}$, is different from $\mathbf{X}_{768}$, the mutational flow is not sufficiently strong to couple $\mathbf{X}_0$, $\mathbf{X}_4$, and $\mathbf{X}_{68}$, and transitions between different quasispecies are observed (Fig. 5.14). The lower diagram ($s = 919$) shows the typical fitness distribution for a strong quasispecies: The fittest two-error mutant, $\mathbf{X}_{516}$, has its fittest nearest neighbor, $\mathbf{X}_4$, in the one-error class Γ_1 and it coincides with the fittest one-error mutant. Accordingly, the three sequences ($\mathbf{X}_0$, $\mathbf{X}_4$, and $\mathbf{X}_{516}$) are strongly coupled by mutational flow and a strong quasispecies is observed (Fig. 5.15).

sequence space. Any sequence in the two-error class, for example, has two nearest neighbors in the one-error class but $n-2$ nearest neighbors in the three-error class. To be a candidate for a strong quasispecies requires that – against probabilities – the fittest sequence in the one-error neighborhood of $\mathbf{X}_{m(2)}$ lies in the one-error class: $(f_{(\mathbf{X}_{m(2)})_{m(1)}})_{\max}$ with $(\mathbf{X}_{m(2)})_{m(1)} \in \Gamma_1$ and preferentially, this is the fittest one-error sequence, $(\mathbf{X}_{m(2)})_{m(1)} \equiv \mathbf{X}_{m(1)}$. Since all mutation rates between nearest neighbor sequences in neighboring classes are the same – $(1-p)^{n-1}p$ within the uniform error model – the strength of mutational flow is dependent only on the fitness values, and the way in which the three sequences were determined guarantees optimality of the flow: If such a three-membered cluster was found it is the one with the highest internal mutational flow for a given landscape. Fig. 5.16 (lower picture, $s = 919$) shows an example where such three sequences form a strongly coupled cluster. There is always a fourth sequence – here $\mathbf{X}_{512}$ – belonging to the cluster but it may play no major role because of low fitness. The heuristic presented was applied to 21 fitness landscapes with different random scatter and three strong quasispecies (s =401, 577, and 919) were observed. How many would be expected by combinatorial arguments? The probability for a sequence in Γ_2 to have a neighbor in Γ_1 is $2/10 = 0.2$ and, since the sequence $\mathbf{X}_{m(1)}$ is fittest in Γ_1 and hence also in the one-error neighborhood of $\mathbf{X}_{m(2)}$, this is also the probability for finding a strong quasispecies. The sample that has been investigated in this study comprised 21 landscapes an hence we expect to encounter $21/5 = 4.2$ cases, which is – with respect to the small sample size – in agreement with the three cases that we found.

The suggestion put forward in the heuristic mentioned above – a cluster of sequences coupled by mutational flow that is stronger within the group than to the rest of sequence space because of frequent mutations and high fitness values – will now be analyzed and tested through the application of zero backflow approximation. Instead of a single master sequence we consider a *master cluster* of sequences and then proceed in full analogy to subsection 4.3.3 by applying zero mutational backflow from the rest of sequence space to the cluster. In order to be able to deal with a cluster of sequences

in the one-error neighborhood, and search only in class $k = 2$.

we rearrange the value matrix W:

$$\mathrm{W} = \begin{pmatrix} W_{11} & W_{12} & \cdots & W_{1k} & W_{1,k+1} & \cdots & W_{1n} \\ W_{21} & W_{22} & \cdots & W_{2k} & W_{2,k+1} & \cdots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ W_{k1} & W_{k2} & \cdots & W_{kk} & W_{k,k+1} & \cdots & W_{kn} \\ W_{k+1,1} & W_{k+1,2} & \cdots & W_{k+1,k} & W_{k+1,k+1} & \cdots & W_{k+1,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \cdots & W_{nk} & W_{n,k+1} & \cdots & W_{nn} \end{pmatrix} . \quad (5.10)$$

The upper left square part of the matrix W will be denoted by w_m. It represents the core of the quasispecies in the sense of a mutationally coupled master cluster, $\mathcal{C}_m = \{\mathbf{X}_{m1}, \ldots, \mathbf{X}_{mk}\}$, and after neglect of mutational backflow from sequences outside the core we are left with the eigenvalue problem

$$\mathrm{w}_m \, \boldsymbol{\zeta}_{mj} \;=\; \lambda_{mj} \, \boldsymbol{\zeta}_{mj} \,; \; j = 0, \ldots, k-1 \; . \quad (5.11)$$

In the uniform error rate model the elements of the mutation matrix Q are of the form

$$\begin{aligned} Q_{mi,mj} \;&=\; (1-p)^{n-d_{mi,mj}} \, p^{d_{mi,mj}} \;=\; (1-p)^{n-k} \, q_{mi,mj} \quad \text{with} \\ q_{mi,mj} \;&=\; (1-p)^{k-d_{mi,mj}} \, p^{d_{mi,mj}} \end{aligned}$$

Apart from the reduced dimension the eigenvalue problem (5.11) is in complete analogy to the eigenvalue problem in subsection 4.3.1. The common factor $(1-p)^{n-k}$ leaves the eigenvectors unchanged and is a multiplier for the eigenvalues: $\lambda_{mj} \Rightarrow (1-p)^{n-k} \, \lambda_{mj} \, \forall \, j = 0, \ldots, k-1$. Only the largest eigenvalue λ_{m0} and the corresponding eigenvector $\boldsymbol{\zeta}_{m0}$ – with the components $\zeta_i^{(m0)}$ and $\sum_{i=0}^{k} \zeta_i^{(m0)} = 1$ – are important for the discussion of the quasispecies. By the same tokens as in subsection 4.3.3, equation (4.19a), we obtain the stationary solution

$$\begin{aligned} \bar{c}_m^{(0)} \;&=\; \frac{\lambda_{m0} \, (1-p)^{n-k} \; - \; \overline{f}_{-m}}{\overline{f}_m \; - \; \overline{f}_{-m}} \quad \text{with} \\ \bar{x}_{mj}^{(0)} \;&=\; \zeta_j^{(m0)} \, \bar{c}_m^{(0)} \,; \; j = 1, \ldots, k \,, \quad \text{and} \\ \overline{f}_m \;&=\; \sum_{i=1}^{k} \zeta_i^{(m0)} f_i \quad \text{and} \quad \overline{f}_{-m} \;=\; \sum_{i=k+1}^{n} \bar{x}_i f_i \Big/ \sum_{i=k+1}^{n} \bar{x}_i \; . \end{aligned} \quad (5.12)$$

The calculations of the concentrations of the sequences not belonging to the master core is straightforward but more involved than in the case of a single

Table 5.2: **Strong quasispecies.** Shown are error thresholds and level crossing values for three cases of strong quasispecies.

Random seeds	$s = 401$		$s = 577$		$s = 919$	
	j	fitness	j	fitness	j	fitness
Core sequences	**0**	1.1000	**0**	1.1000	**0**	1.1000
	64	1.0981	**64**	1.0951	**4**	1.0966
	16	1.0772	**256**	1.0894	**512**	0.9296
	80	1.0987	**320**	1.0999	**516**	1.0970
	j	$p_{(1/100)}$	j	$p_{(1/100)}$	j	$p_{(1/100)}$
Level crossing, $d = 0$	**0**	0.01396	**0**	**0.01410**	**0**	0.01320
	64	**0.01406**	**64**	0.01402	**4**	**0.01348**
	16	0.01318	**256**	0.01377	**512**	0.00828
	80	0.01389	**320**	0.01410	**516**	0.01304
Level crossing, $d = 1$	**0**	**0.008443**	**0**	**0.006921** [a]	**0**	0.007876
	64	0.008359	**64**	0.006481	**4**	**0.008385**
	16	0.007003	**256**	0.006440	**512**	- - - [b]
	80	0.007876	**320**	0.006733	**516**	0.007476
Error threshold (p_{cr})	0.01134		0.01145		0.01087	

[a] The quasispecies with $s = 577$ shows a small smooth transition just above the error threshold. The following three sequences have the same or higher level crossing values: $p_{(1/100)}(\mathbf{X}_{899}) = 0.008026$, $p_{(1/100)}(\mathbf{X}_{931}) = 0.007186$, and $p_{(1/100)}(\mathbf{X}_{962}) = 0.006842$.

[b] The stationary concentration $\bar{x}_{512}(p)$ never exceeds nor reaches the value 0.01.

master sequence. We dispense here from details because we shall not make use of the corresponding expressions. In the forthcoming examples we shall apply a modified single peak landscape where all sequences except those in the master core have the same fitness values f and then the equation $\overline{f}_{-m} = f$ is trivially fulfilled.

For the purpose of illustration of the analysis of sequence clustering in strong quasispecies full numerical computations are compared with the zero mutational backflow approximation for the four membered cluster on the fitness landscape $\mathcal{L}(\lambda = 0, s = 919, d = 1.00)$ in Fig. 5.17. Although differences are readily recognized and the agreement between the full calculation and the approximation is not as good as in the case of a single master sequence, the appearance of the cluster is very well reproduced by the zero mutational backflow approximation. In particular, the relative frequencies of the four sequences forming the cluster are reproduced well. In comparison to the full calculation, the critical mutation rate at the error threshold, the point $p = p_{\text{cr}}$ at which the entire quasispecies $\bar{\Upsilon}^{(0)}$ vanishes, appears at a higher p-value than the level crossings of the full calculation. The difference in the critical mutation rates is readily interpreted: The full calculation is based on a landscape with fully developed random scatter whereas the zero mutational backflow calculation compares best with a *four peak landscape* where the four peaks correspond to the members of the cluster ($\mathbf{X}_0$, $\mathbf{X}_4$, $\mathbf{X}_{516}$, $\mathbf{X}_{512}$) and all other sequences have identical fitness values. In order to show that this interpretation is correct the cluster has been implemented on a single peak landscape ($d = 0$) with the same fitness values ($f_0 = 1.1$ and $f = 1.0$) and the error threshold on this landscape is shifted slightly to higher values of the mutation rate parameter p. The agreement with the zero mutational backflow approximation is remarkably good. This agreement can be taken as a strong indication that the interpretation of strong quasispecies being a result of the formation of mutationally linked clusters of sequences within the population.

Three strong quasispecies with the values $s = 401$, 577, and 919 were found among the 21 landscapes studied here. The most important computed data are summarized in table 5.2. Like in the single master case on the single peak landscape the level crossing values $p_{(1/100)}$ occur at higher mutation rates than the error threshold (see Fig. 5.10). The shift in the strong quasispecies is about $\Delta p = 0.0265$ somewhat larger than that for the single master lying at $\Delta p = 0.00226$. In the single master case the error threshold was calculated to be $p_{\text{cr}} = 0.094875$ whereas here it is shifted to higher p-values by about

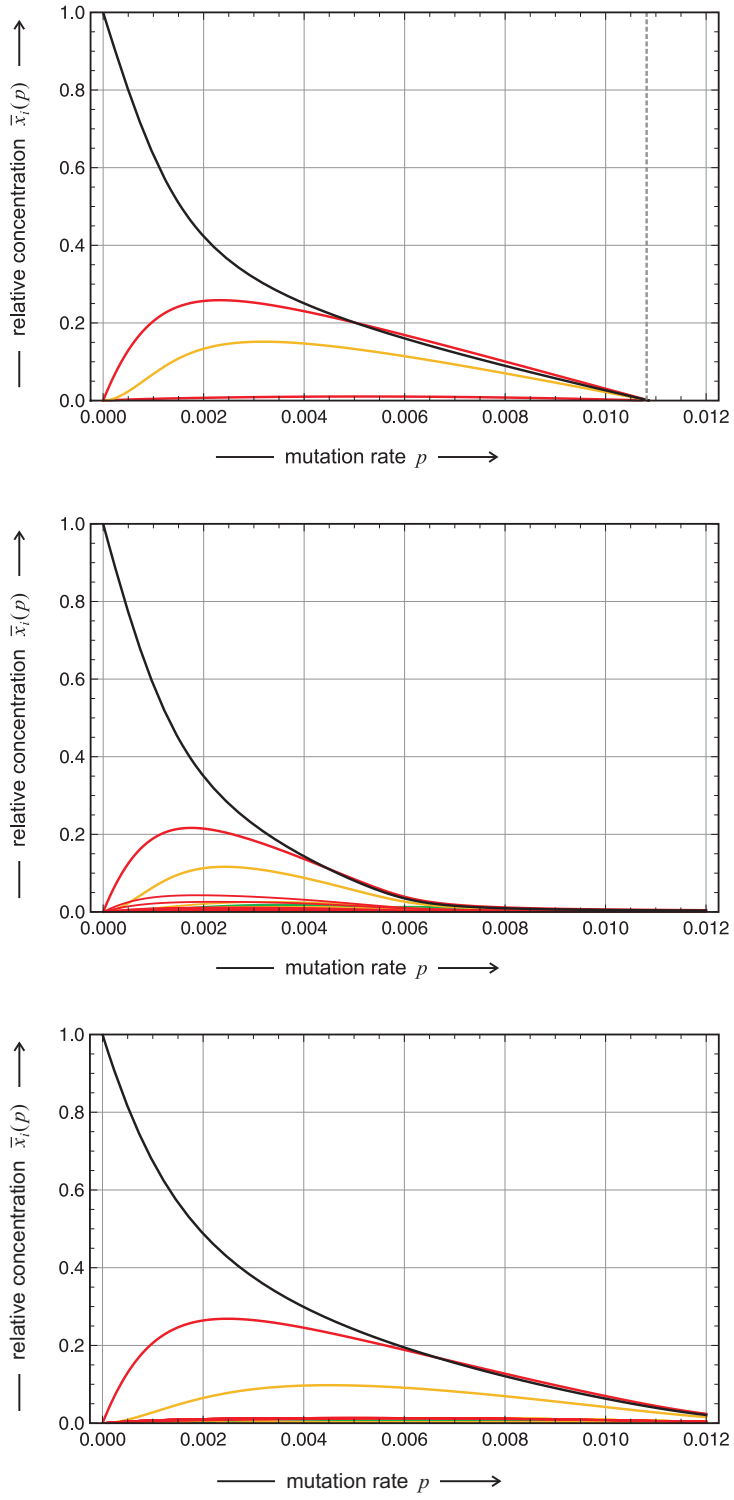


Figure 5.17: **Zero backflow approximation for a quasispecies on a *realistic* model landscape.** The landscape characteristic is $\mathcal{L}(\lambda = 0, d = 1.00, s = 919)$. Shown are the stationary concentrations $\bar{x}_j(p)^{(0)}$ $(j = 1, 2, 3, 4)$ for the cluster obtained through zero mutational backflow (upper plot), the results of the full numerical computation (middle plot), and of a full numerical computation where the cluster was implemented on the single peak landscape (lower plot, $d = 0$). Other parameters: $l = 10$, $f_0 = 1.1$, and $f = 1.0$.

$\Delta p = 0.0046$. The interpretation is straightforward: The core taken together has a higher effective fitness than a single master and this is reflected by the shift to higher mutation rates. This shift is smallest in case of the core with $s = 919$ being the in agreement with a particularly small fitness value of one of the two class 1 mutant ($f_{512} = 0.9296$). In fact, the core in this case consists practically of three sequences only: $\mathbf{X}_0$, $\mathbf{X}_4$, and $\mathbf{X}_{516}$. In the computation with fully developed scatter ($d = 1$) for the strong quasispecies with $s = 577$ we observe $p_{(1/100)}$-values that are smaller than in the other two cases. Again the explanation is straightforward: There is a small and smooth transition at a p_{tr} value just below the error threshold and the stationary concentration of the master beyond the transition is higher than that of the dominating sequence in the core, $\bar{x}_{899} > \bar{x}_0$, and the $p_{(1/100)}$-value for the sequence $\mathbf{X}_{899}$ is indeed higher, $p_{(1/100)}(\mathbf{X}_{899}) = 0.008026$.

The mutation rate at which the last stationary concentration crosses the value 1/100 shows some scatter: For the twenty one random landscapes that were investigated here it amounts to $p_{(1/100)}(\mathbf{X}_{\mathrm{last}}) = 0.00812 \pm 0.00071$. Interestingly the values for strong quasispecies lie close together at $p_{(1/100)}(\mathbf{X}_{\mathrm{last}}) = 0.0084$. The observed scatter in the level crossing of the concentration of $\mathbf{X}_{\mathrm{last}}$ is definitely smaller than that found for the master sequence ($p_{(1/100)}(\mathbf{X}_0)$ in Fig. 5.12), which is an obvious result since $\bar{x}_0$ decays to small values at transitions that occur before the error threshold, at values $p_{\mathrm{tr}} < p_{\mathrm{cr}}$.

5.3 Dynamics on realistic rugged and neutral landscapes

The second property of realistic fitness landscapes mentioned in section 5.2 is *neutrality* and the challenge is to implement it together with ruggedness. In order to be able to handle both features together we conceived a two parameter landscape: (i) the random scatter is denoted by d as before and (ii) a degree of neutrality λ is introduced. The value $\lambda = 0$ means absence of neutrality and $\lambda = 1$ describes the completely flat landscape in the sense of Motoo Kimura's *neutral evolution* [174]. The result of the theory of neutral evolution that is most relevant here concerns *random selection*: Although fitness differences are absent, one randomly chosen sequence is selected by means of the stochastic replication mechanism, $\mathbf{X} \rightarrow 2\,\mathbf{X}$ and $\mathbf{X} \rightarrow \oslash$. For most of the time the randomly replicating population consists of a dominant genotype and a number of neutral variants at low concentration.

An important issue of the landscape approach is the random positioning of neutral master sequences in sequence space, which is achieved by means of the same random number generator that is used to compute the random scatter of the other fitness values obtained from pseudorandom numbers with a uniform distribution in the interval $0 \leq \eta \leq 1$:

$$f(\mathbf{X}_j) = \begin{cases} f_0 & \text{if } \; j = 0 \; , \\ f_0 & \text{if } \; \eta_j^{(s)} \geq 1 - \lambda \; , \\ f + \frac{2d}{1-\lambda}(f_0 - f)\left(\eta_j^{(s)} - 0.5\right) & \text{if } \; \eta_j^{(s)} < 1 - \lambda \; , \\ & j = 1, \ldots, \kappa^l; \; j \neq m \; . \end{cases} \tag{5.13}$$

The rugged and neutral fitness landscape (5.13) is the complete analogue to the rugged fitness landscape (5.6) under the condition that several master sequences exist, which have the same highest fitness values f_0. The fraction of neutral mutants is determined by the fraction random numbers, which fall into the range $1 - \lambda < \eta \leq 1$, apart from statistical fluctuations this fraction is λ. At small values of the degree of neutrality λ isolated peaks of highest fitness f_0 will appear in sequence space. Increasing λ will result in the formation of clusters of sequences of highest fitness. Connecting all

fittest sequences of Hamming distance $d_{\mathrm{H}} = 1$ by an edge results in a graph that has been characterized as *neutral network* [238, 239]. Neutral networks were originally conceived as a tool to model, analyze, and understand the mapping of RNA sequences into secondary structures [126, 127, 252]. The neutral network in RNA sequence-structure mappings is the preimage of a given structure in sequence space and these networks can be approximated in zeroth order by random graphs [78, 79]. Whereas neutral networks in RNA sequence-structure mappings are characterized by a relatively high degree of neutrality around $\lambda \approx 0.3$ and sequence space percolation is an important phenomenon, we shall be dealing here with much lower λ-values.

5.3.1 Small neutral clusters

The two smallest clusters of fittest sequences have Hamming distances $d_{\mathrm{H}} = 1$ and $d_{\mathrm{H}} = 2$ (Fig. 5.18). In the former case we are dealing with a minimal neutral neutral network, in the latter case the Hamming distance two sequences are coupled through two intermediate sequences similarly as in the core of strong quasispecies. An exact mathematical analysis for both cases is possible in the limit of vanishing mutation rates, $\lim p \to 0$ [256], led to results that differ from Kimura's neutral theory:

$$\lim_{p\to 0} \bar{x}_{\mathrm{I}} = \frac{1}{2}\,,\ \lim_{p\to 0} \bar{x}_{\mathrm{II}} = \frac{1}{2} \qquad \text{for } d_{\mathrm{H}}(\mathbf{X}_{\mathrm{I}}, \mathbf{X}_{\mathrm{II}}) = 1\,, \tag{5.14a}$$

$$\lim_{p\to 0} \bar{x}_{\mathrm{I}} = \frac{\alpha}{1+\alpha}\,,\ \lim_{p\to 0} \bar{x}_{\mathrm{II}} = \frac{1}{1+\alpha} \qquad \text{for } d_{\mathrm{H}}(\mathbf{X}_{\mathrm{I}}, \mathbf{X}_{\mathrm{II}}) = 2\,, \tag{5.14b}$$

$$\lim_{p\to 0} \bar{x}_{\mathrm{I}} = 1\,,\ \lim_{p\to 0} \bar{x}_{\mathrm{II}} = 0 \quad \text{or} \quad \lim_{p\to 0} \bar{x}_{\mathrm{I}} = 0\,,\ \lim_{p\to 0} \bar{x}_{\mathrm{II}} = 1\,, \quad \text{for } d_{\mathrm{H}}(\mathbf{X}_{\mathrm{I}}, \mathbf{X}_{\mathrm{II}}) \geq 3\,. \tag{5.14c}$$

If the two neutral fittest sequences, $\mathbf{X}_{\mathrm{I}}$ and $\mathbf{X}_{\mathrm{II}}$, are nearest neighbors in sequence space, $d_{\mathrm{H}}(\mathbf{X}_{\mathrm{I}}, \mathbf{X}_{\mathrm{II}}) = 1$, they are present at equal concentrations in the quasispecies in the low mutation rate limit, in case they are next nearest neighbors in sequence space, $d_{\mathrm{H}}(\mathbf{X}_{\mathrm{I}}, \mathbf{X}_{\mathrm{II}}) = 2$, they are observed at some ratio α, and in both cases none of the two sequences vanishes. Only for

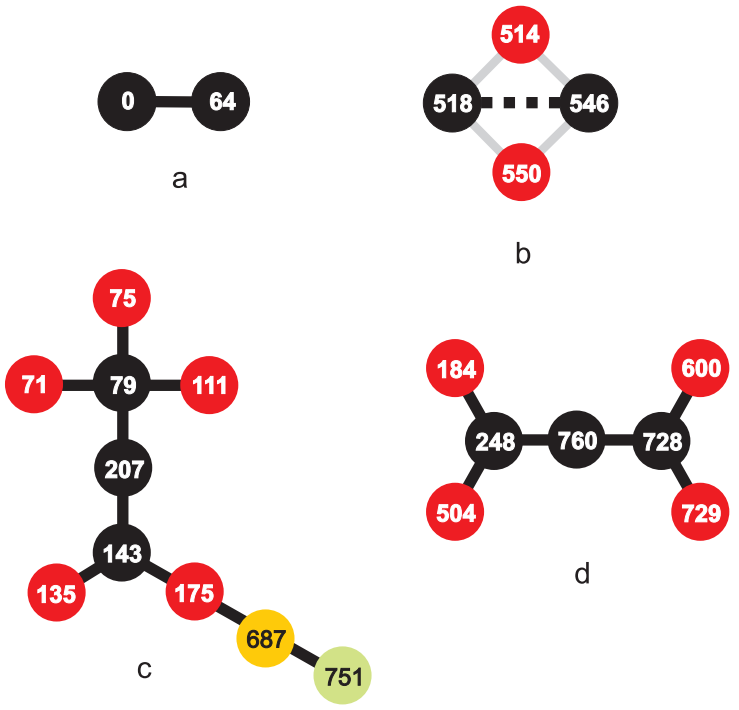


Figure 5.18: **Neutral networks in quasispecies.** The sketch presents four special cases that were observed on rugged neutral landscapes defined in equation (5.13). Part **a** shows the smallest possible network consisting of two sequences of Hamming distance $d_{\mathrm{H}} = 1$ observed with $s = 367$ and $\lambda = 0.01$. Part **b** contains two sequences of Hamming distance $d_{\mathrm{H}} = 2$, which are coupled through two $d_{\mathrm{H}} = 1$ sequences; it was found with $s = 877$ and $\lambda = 0.01$. The neutral network in part **c** has a core of three sequences, surrounded by five one-error mutants, one of them having a chain of two further mutants attached to it; the parameters of the landscape are $s = 367$ and $\lambda = 0.1$. Part **d** eventually shows a symmetric network with three core sequences and four one-error mutants attached to it, observed with $s = 229$ and $\lambda = 0.1$. Choice of further parameters: $n = 10$, $f_0 = 1.1$, $f = 1.0$, and $d = 0.5$. Color code: core sequences in black, one-error mutants in red, two-error mutants in yellow, and three-error mutants in green.

Hamming distances $d_{\mathrm{H}}(\mathbf{X}_{\mathrm{I}}, \mathbf{X}_{\mathrm{II}}) \geq 3$ Kimura's scenario of random selection occurs. It is important to stress a difference between the two scenarios, the deterministic ODE approach leading to clusters of neutral sequences and the random selection phenomenon of Motoo Kimura: In the quasispecies we have strong mutational flow within the cluster of neutral sequences – which is not substantially different from the flow within the non-neutral clusters

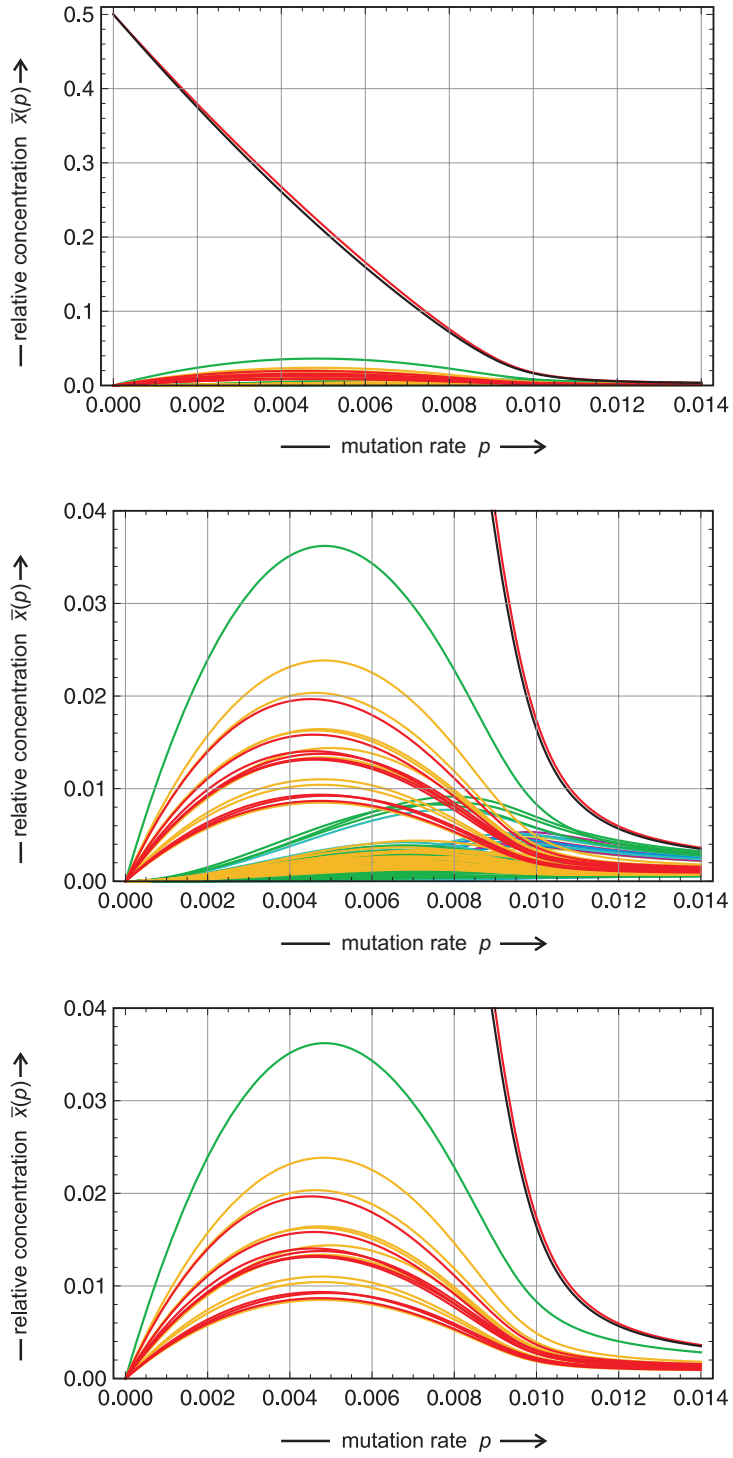


Figure 5.19: **Cluster on a weakly neutral rugged model landscape with $s = 367$.** The plot in the middle is an enlargement of the topmost plot. in the bottom plot only the curves of the dominant cluster, consisting of the two master sequences, $\mathbf{X}_0$ and $\mathbf{X}_{64}$, their one-error neighborhoods, and the third fittest neutral sequence $\mathbf{X}_{324}$, are shown. Further parameters: $n = 10$, $f_0 = 1.1$, $f = 1.0$, $\lambda = 0.01$, $d = 0.5$. Color code see appendix.

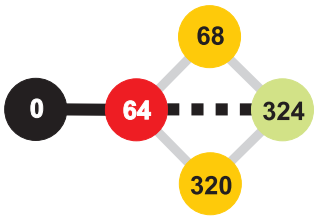


Figure 5.20: **A small neutral cluster in a quasispecies.** The color code from the appendix that is different from Fig. 5.18 is used: $\mathbf{X}_0$ black, the one-error mutant $\mathbf{X}_{64}$ red, the two-error mutants in yellow, and the fittest three-error mutant $\mathbf{X}_{324}$ in green.

in subsection 5.2.3 – and this flow outweighs fluctuations. In the random replication scenario mutations don't occur and the only drive for change in particle numbers is random fluctuations. For Hamming distances d_{H} of three and more the mutational flow is too weak to counteract random drift.

An important issue of the quasispecies dynamics of closely related neutral sequences ($d_{\mathrm{H}} = 1$ and $d_{\mathrm{H}} = 2$) is the fact that mutational backflow between two or more master sequences leads to a situation where population dynamics can be described already on the deterministic level in contrast to Kimura's model of random replication where *random selection* is a phenomenon that is exclusively observable in populations of finite size. The difference between both approaches is caused by mutational flow between neutral sequences: the coupling of neutral sequences of Hamming distance one or two forms a strongly coupled cluster of sequences of the kind discussed also in the non-neutral case and this cluster behaves like a single entity competing with its mutant cloud.

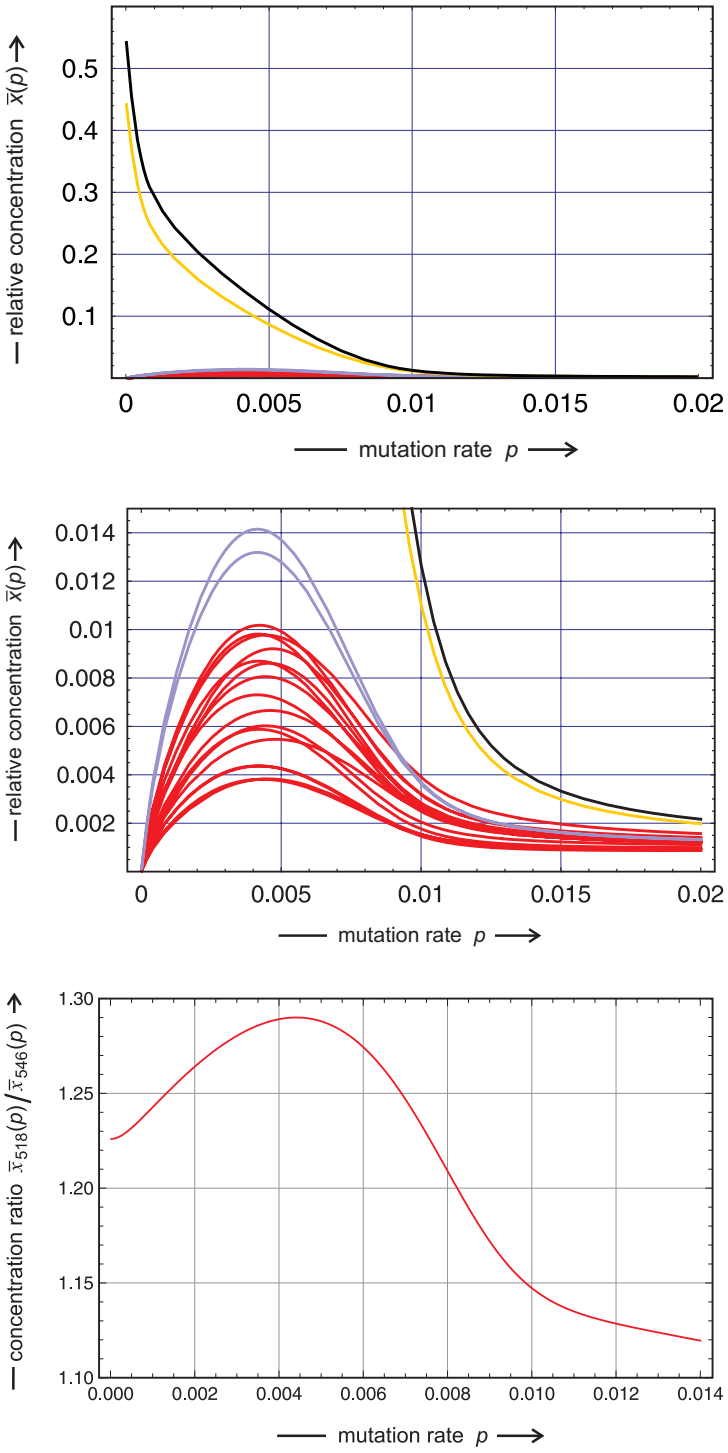


Figure 5.21: **Quasispecies on weakly neutral rugged model landscape with $s = 877$.** The topmost part of the figure refers to the landscape with $s = 877$ and presents the solution curves for the master pair, $\{\mathbf{X}_{518}, \mathbf{X}_{546}\}$ and their one-error mutants. The plot in the middle is an enlargement and highlights the curves for the two intermediate sequences $\mathbf{X}_{514}$ and $\mathbf{X}_{550}$ in pastel blue. The plot at the bottom shows the ratio between the stationary concentrations of the two master sequences, $\alpha(p)$. Further parameters: $n = 10$, $f_0 = 1.1$, $f = 1.0$, $\lambda = 0.01$, $d = 0.5$. Color code see appendix.

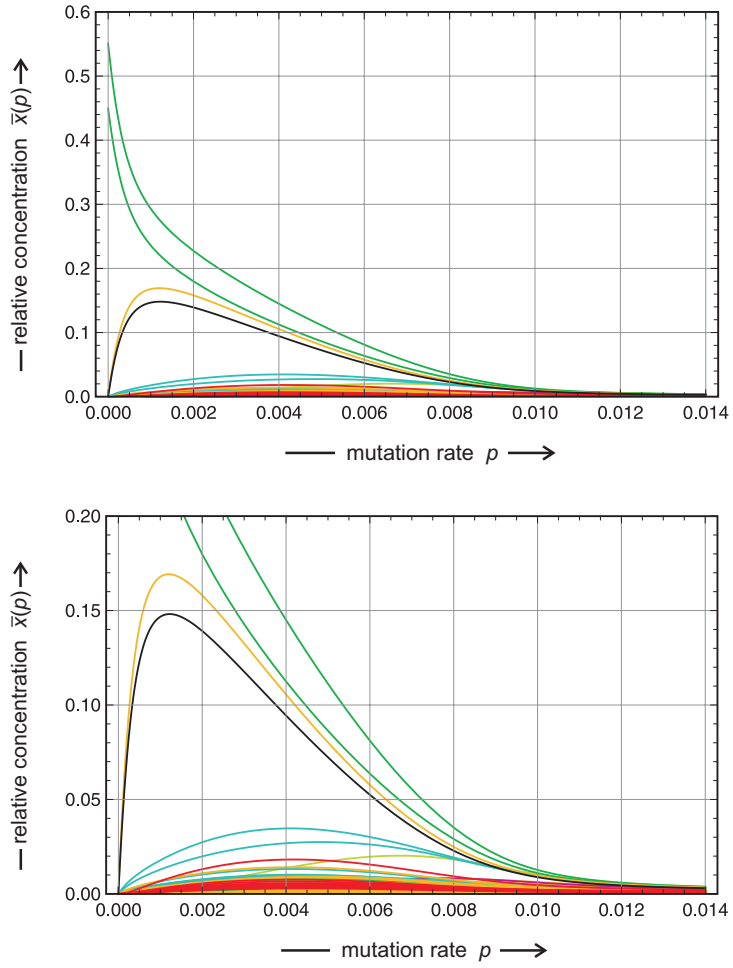


Figure 5.22: **Quasispecies on weakly neutral rugged model landscape with $s = 877$.** The topmost part of the figure refers to the landscape with $s = 877$ and presents all solution curves Further parameters: $n = 10$, $f_0 = 1.1$, $f = 1.0$, $\lambda = 0.01$, $d = 0.5$. Color code see appendix.

The question now is whether or not the exact results derived for $\lim p \to 0$ are of more general validity. In order to find an answer numerical computations of quasispecies as functions of the mutation rate p were performed. Random landscapes with a degree of neutrality of $\lambda = 0.01$ were searched and indeed the desired small networks with distances $d_{\mathrm{H}} = 1$ and one for $d_{\mathrm{H}} = 2$ between the master sequences were found for $s = 367$ and $s = 877$ (Fig. 5.18, parts **a** and **b**, respectively). Figs. 5.19 and 5.21 show the solutions

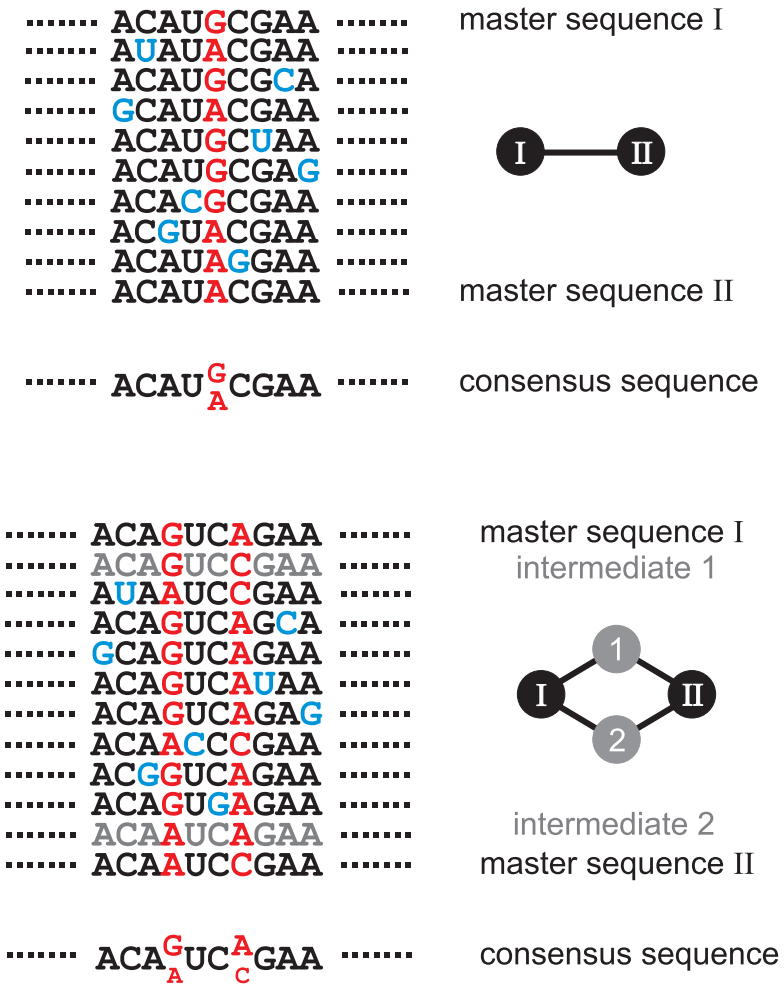


Figure 5.23: **Quasispecies with two neutral master sequences.** The sketch contains a set of sequences in order to demonstrate the role of neutrality in the determination of the consensus sequence of a population. Two fittest sequences with Hamming distance $d_{\mathrm{H}} = 1$ (upper picture) lead to an ambiguity at one position. Mutations at other positions are wiped out by statistics whereas the one-to-one ratio of the two mater sequences leads to a 50/50 ratio of two nucleobases at the position of the mutation. The lower picture refers to two master sequences with Hamming distance $d_{\mathrm{H}} = 2$: Ambiguities are obtained at two positions and the ratios of the nucleobases are given approximately by the value of α in equation(5.14b). The two intermediates are present in small stationary concentrations only but they are, nevertheless, more frequent than the other one-error mutants (see Fig. 5.21).

curves $\bar{x}_j(p)$ for the two examples of small neutral clusters. The concentration ratios of the two fittest sequences fulfil the predictions of the analytical approach in the limit of small mutation rates, $\lim p \to 0$: The ratio for $\mathbf{X}_0$ and $\mathbf{X}_{64}$ in the Hamming distance one case, $\bar{x}_0(0)/\bar{x}_{64}(0) = 1$, and some finite ratio $\bar{x}_{518}(0)/\bar{x}_{546}(0) = \alpha = 1.2259$ in the Hamming distance two case, respectively.

Fig. 5.19 illustrates the dependence of the quasispecies formed by two master sequences of Hamming distance $d_{\mathrm{H}} = 1$ on the mutation rate p. The extrapolation of the exact result, $\bar{x}_0/\bar{x}_{64} = 1$ to nonzero mutation rates turns out to be successful: Indeed, the red curve behind the black curve is hardly to be seen in the topmost plot as well as in the enlargement (middle plot). A precise calculation of this ratio shows a slight increase until at $\hat{p} = 0.009405$ a maximum of $\bar{x}_0(\hat{p})/\bar{x}_{64}(\hat{p}) = 1.0610$ is reached. Then the ratio decreases and apparently becomes unity again at $\tilde{p} = 0.5$. The plot of all stationary concentrations in the quasispecies belonging to the network **a** in Fig. 5.18 shows an interesting detail: The non-master sequence with the highest concentration, $\mathbf{X}_{324}$, does not belong to the combined one-error neighborhood of the two master sequences but lies at Hamming distance $d_{\mathrm{H}} = 2$ and $d_{\mathrm{H}} = 3$ from the two masters, $\mathbf{X}_0$ and $\mathbf{X}_{64}$, respectively. The explanation follows straightforwardly from an inspection of the fitness landscape. Sequence $\mathbf{X}_{324}$ belongs to the class of fittest neutral sequences but it is not coupled by an edge to the dominant network (Fig. 5.20). Instead it forms an Hamming distance two cluster together with $\mathbf{X}_{64}$ with $\mathbf{X}_{68}$ and $\mathbf{X}_{320}$ being the intermediates. The plot at the bottom of Fig. 5.19 contains the curves of the stationary concentrations of the Hamming distance one master pair (black and red) and their complete one-error neighborhood (red and yellow, respectively) together with that of the neutral sequence $\mathbf{X}_{324}$. It is interesting that all curves of the neighborhood sequences and the curve of $\mathbf{X}_{324}$ have their maxima at almost the same position near $p \approx 0.05$ whereas the maxima of all other curves (comparison with the middle plot of Fig. 5.20) are shifted towards higher p-values. An error threshold expressed by means of $p_{1/100}$-values occurs at somewhat higher mutation rates than in the case of a single master sequence: $p_{1/100}(\mathbf{X}_0) = 0.01073$ and $p_{1/100}(\mathbf{X}_{\mathrm{last}}) = 0.01082$ with $\mathbf{X}_{\mathrm{last}} \equiv \mathbf{X}_{64}$ compared

to $p_{1/100}(\mathbf{X}_0) = 0.01065$ in the non-neutral case. As expected the Hamming distance one pair is equivalent to a master that is slightly stronger than a single sequence. Indeed, the fitness value of $\mathbf{X}_{64}$ is raised from $f_{64} = 1.04923$ to $f_{64} = 1.1$ on the landscape with neutrality.

An isolated cluster with a distance $d_{\mathrm{H}} = 2$ between the two master sequences, $\{\mathbf{X}_{518}, \mathbf{X}_{546}\}$, has been observed on the rugged neutral landscape with $\lambda = 0.01$ and $s = 877$. In the limit $p \to 0$ the two fittest neutral sequences $\mathbf{X}_{518}$ and $\mathbf{X}_{546}$ are present at the stationary concentrations $\bar{x}_{518} = 0.5507$ and $\bar{x}_{546} = 0.4493$, respectively, and their ratio is $\bar{x}_{518}(0) \,/\, \bar{x}_{546}(0) = \alpha(0) = 1.2259$. Both stationary concentrations decrease with increasing p-values, the ratio increases at first but then decreases and approaches the value one corresponding to the uniform distribution: $\lim_{p\to\tilde{p}=1/2} \alpha(p) = 1$. The function $\alpha(p)$ passes a (local) maximum of $\alpha(p) = 1.29$ at $p = 0.00441$ (see Fig. 5.21, plot at the bottom). The plot in the middle of the figure demonstrates that the two sequences lying in between the master pair, $\mathbf{X}_{514}$ and $\mathbf{X}_{550}$, appear at higher concentrations than the rest of the one-error cloud.[13] It is interesting to note that the landscape $\mathcal{L}(\lambda = 0.01, s = 877)$ sustains another Hamming distance two pair of fittest sequences $\{\mathbf{X}_0, \mathbf{X}_{132}\}$ with the intermediates $\mathbf{X}_4$ and $\mathbf{X}_{128}$. This second cluster is in competition with the first cluster as shown in Fig. 5.22 and gains in concentration with increasing mutation rates p, passes through a maximum and then decays through an error threshold to the uniform distribution at $p = \tilde{p}$. The position of the error threshold again is estimated by means of the $p_{1/100}$-values and one finds $p_{1/100}(\mathbf{X}_{518}) = 0.01053$ and $p_{1/100}(\mathbf{X}_{546}) = 0.01022$ with $\mathbf{X}_{518} \equiv \mathbf{X}_{\mathrm{last}}$. On this neutral landscape the corresponding non-neutral master sequence is more stable than the cluster as expressed by $p_{1/100}(\mathbf{X}_0) = 0.01075$. This fact is difficult to interpret, because the original master $\mathbf{X}_0$ is not member of the cluster, which accordingly is situated in another part of sequences space with different fitness values of the neighboring sequences.

[13] In the case shown here, the two intermediate sequences have very similar fitness values, $f_{514} = 1.017$ and $f_{550} = 1.012$. Large fitness differences can outweigh the advantage caused by the mutation flow from both master sequences.

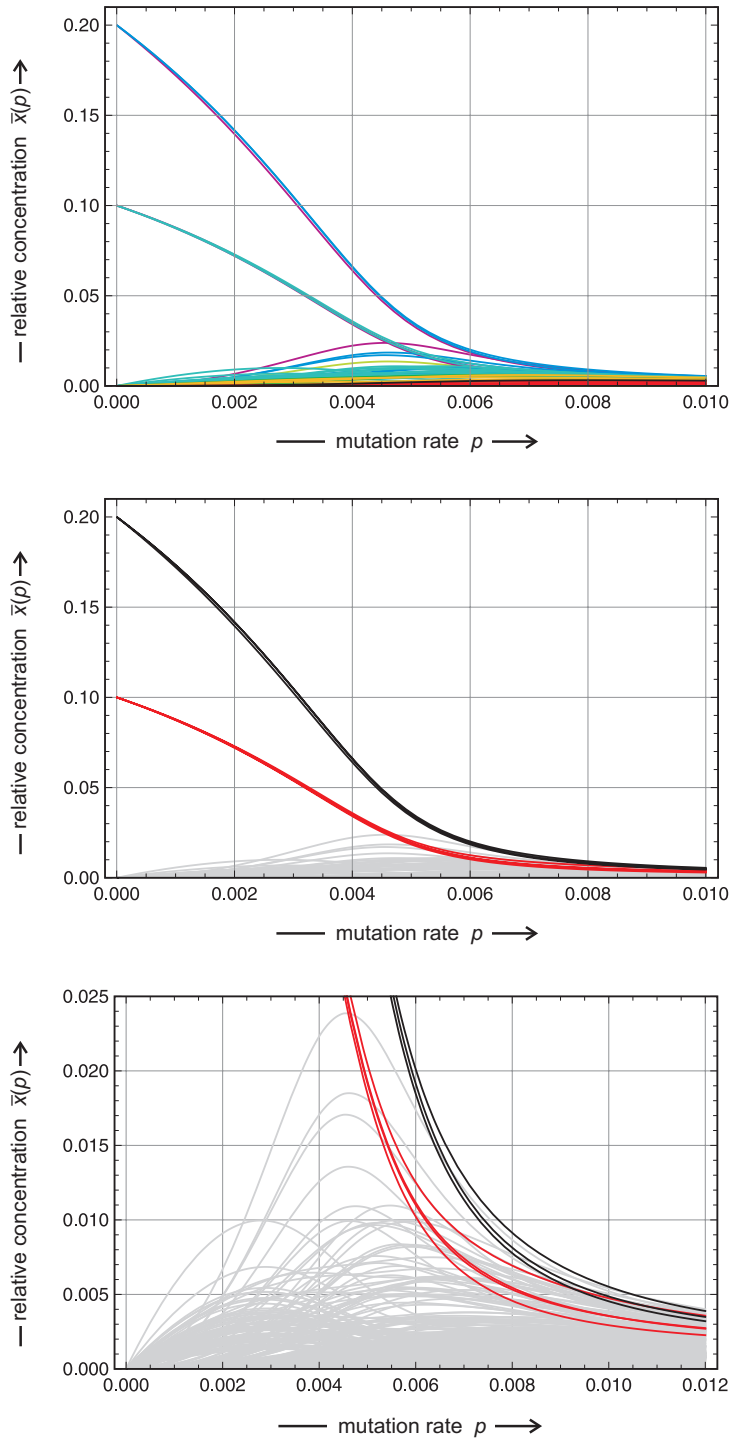


Figure 5.24: **Quasispecies on rugged neutral model landscapes I.** Shown are the stationary concentrations for the landscape $\mathcal{L}(\lambda = 0.1, d = 0.5, s = 229)$. The topmost plot is drawn with the color code of the appendix, the plot in the middle applies the color code of the neutral network in Fig. 5.18 **d** with the curves of sequences not belonging to the net in grey. The plot at the bottom is an enlargement of the plot in the middle. Other parameters: $l = 10$, $f_0 = 1.1$, and $f = 1.0$.

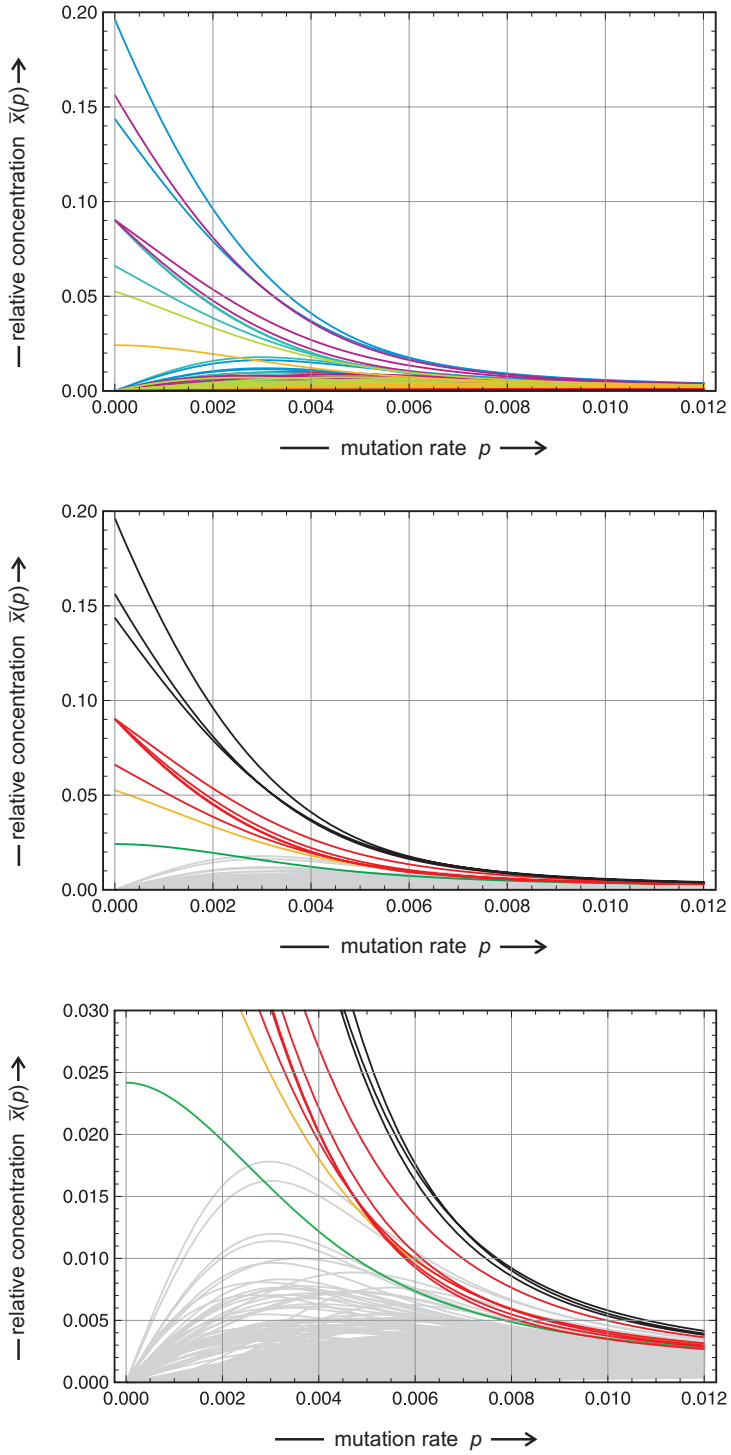


Figure 5.25: **Quasispecies on rugged neutral model landscapes II.** Shown are the stationary concentrations for the landscape $\mathcal{L}(\lambda = 0.1, d = 0.5, s = 367)$. The topmost plot is drawn with the color code of the appendix, the plot in the middle applies the color code of the neutral network in Fig. 5.18 **c** with the curves of sequences not belonging to the net in grey. The plot at the bottom is an enlargement of the plot in the middle. Other parameters: $l = 10$, $f_0 = 1.1$, and $f = 1.0$.

Eventually we consider a simple practical consequence of the existence of fittest neutral pairs of Hamming distance $d_{\mathrm{H}} = 1$ and $d_{\mathrm{H}} = 2$ for the sequence analysis in populations. Despite vast sequence heterogeneity [55], in particular of virus populations, average or consensus sequences are fairly insensitive to individual mutations provided the population size is sufficiently large. Individual deviations in mutant sequences cancel through averaging in population with single master sequences. This will not be the case in the presence of neutral variants. In the 50/50 mixture of two master sequences with mutant clouds surrounding both the sequence difference between the masters is not going to cancel by averaging and ambiguities remain. Considering now the two cases discussed here: (i) two master sequences at Hamming distance one present at equal concentrations and (ii) two master sequences at Hamming distance two present at a concentration ratio α, we expect to find sequence averages as sketched in Fig. 5.23. In the former case a 50/50 mixtures of two nucleotides is expected to occur at one position on the sequence, and in the latter case two positions will show nucleobase ambiguities with the ratio α.

5.3.2 Medium size neutral clusters

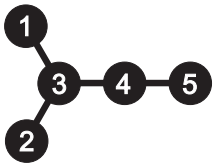
An increase in the degree of neutrality λ will results in the appearance of larger neutral networks that are scattered all over sequence space. We start here by the introduction of the adjacency matrix as an appropriate reference state of neutral networks and the discuss two examples of more complex neutral networks that are observed in form of quasispecies on landscapes $\mathcal{L}_n(\lambda = 0.1, d = 0.5, s)$.

The adjacency matrix of a graph contains an entry one at every off diagonal element that corresponds to an edge in the graph. We have, for example,

the adjacency matrix A

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

for the graph:



Now we consider a neutral network corresponding to this graph and obtain for the mutation matrix Q:

$$Q = \begin{pmatrix} (1-p)^n & (1-p)^{n-2}p^2 & (1-p)^{n-1}p & (1-p)^{n-2}p^2 & (1-p)^{n-3}p^3 \\ (1-p)^{n-2}p^2 & (1-p)^n & (1-p)^{n-1}p & (1-p)^{n-2}p^2 & (1-p)^{n-3}p^3 \\ (1-p)^{n-1}p & (1-p)^{n-1}p & (1-p)^n & (1-p)^{n-1}p & (1-p)^{n-2}p^2 \\ (1-p)^{n-2}p^2 & (1-p)^{n-2}p^2 & (1-p)^{n-1}p & (1-p)^n & (1-p)^{n-1}p \\ (1-p)^{n-3}p^3 & (1-p)^{n-3}p^3 & (1-p)^{n-2}p^2 & (1-p)^{n-1}p & (1-p)^n \end{pmatrix} .$$

In the limit of small mutation rates we neglect all powers $f(p) \in o(p)$ and after multiplication with the fitness matrix $\mathrm{F} = f_0 \cdot \mathbb{I}$, where $\mathbb{I}$ is the identity or unit matrix, the result is the value matrix in the zeroth order approximation

$$W^{(0)} = f_0\, p^{n-1} \cdot \begin{pmatrix} 1-p & 0 & p & 0 & 0 \\ 0 & 1-p & p & 0 & 0 \\ p & p & 1-p & p & 0 \\ 0 & 0 & p & 1-p & p \\ 0 & 0 & 0 & p & 1-p \end{pmatrix} .$$

Since neither the addition of a constant to the diagonal elements nor the multiplication by a common factor changes eigenvectors, the adjacency matrix and the matrix $W^{(0)}$ have identical eigenvectors. Accordingly, the adjacency matrix of the neutral network is the appropriate reference for quasispecies

Table 5.3: **Neutral networks and adjacency matrix.** The largest eigenvector of the adjacency matrix of the neutral network in Fig. 5.18 **c** is compared with the quasispecies calculated for a small value of the mutation rate p.

Class	No.	j	$\boldsymbol{\zeta}_0(\mathbf{X}_j)$ [a]	$\boldsymbol{\zeta}_0^{(0)}(\mathbf{X}_j)$ [b]
Core	1	**79**	0.196220	0.196281
	2	**207**	0.156240	0.156284
	3	**143**	0.143652	0.143688
Class 1	4	**175**	0.090212	0.090231
	5	**71**	0.090206	0.090231
	6	**75**	0.090212	0.090231
	7	**111**	0.090206	0.090231
	8	**135**	0.066039	0.066053
Class 2	9	**687**	0.052585	0.052934
Class 3	10	**751**	0.024177	0.024177

[a] The entries in this column are the components of the quasispecies expressed as the elements of the largest eigenvector of the value matrix W computed with a mutation rate $p = 1 \times 10^{-6}$.

[b] The entries in this column are the components of the largest eigenvector $\boldsymbol{\zeta}_0^{(0)}$ of the adjacency matrix of the graph representing the fittest neutral network on the landscape $\mathcal{L}(\lambda = 0.05, d = 0.5, s = 367)$.

rugged neutral landscapes. Clearly, this has been the case for the small cluster shown in Fig. 5.19 where the dominant eigenvector of the trivial adjacency matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{simply is} \quad \boldsymbol{\zeta}_0^{(0)} \;=\; (\frac{1}{2}, \frac{1}{2})^{\mathrm{t}} \;,$$

and represents also the solution of the mutation selection equation (4.9) for $p = 0$.

In Fig. 5.24 the quasispecies as a function of the mutation rate, $\bar{\Upsilon}(p)$ is shown for the landscape $\mathcal{L}(\lambda = 0.1, d = 0.5, s = 229)$. The neutral network

consists of seven sequences, three of them form a linear inner core and four are attached to it on the periphery (Fig. 5.18 **d**). The dominant eigenvector of the adjacency matrix is of the form

$$\boldsymbol{\zeta}_0^{(0)} = (0.1, 0.1, 0.2, 0.2, 0.2, 0.1, 0.1) = (\bar{x}_{184}, \bar{x}_{504}, \bar{x}_{248}, \bar{x}_{760}, \bar{x}_{728}, \bar{x}_{600}, \bar{x}_{729})^{\mathrm{t}} .$$

The figure shows that the relative concentration within the quasispecies in the sense of three more frequent and four less frequent sequences are perfectly maintained almost up to the error threshold. The level crossing of the three core sequences occurs at: $p_{1/100}(\mathbf{X}_{248}) = 0.007712$, $p_{1/100}(\mathbf{X}_{760}) = 0.007307$, and $p_{1/100}(\mathbf{X}_{728}) = 0.007413$.

The neutral network on the landscape $\mathcal{L}(\lambda = 0.1, d = 0.5, s = 367)$ has a more complicated structure than the symmetric seven membered neutral cluster discussed in the previous paragraph. It contains ten individual sequences and has the form shown in Fig. 5.18 **c**: A core of three sequences is surrounded by five nearest neighbors and has a tail consisting of one Hamming distance two and one Hamming distance three sequence. Also for this more involved topology the low mutation rate limit of the quasispecies, $\lim_{p\to 0} \bar{\Upsilon}(p)$, converges exactly to the largest eigenvector of the adjacency matrix (table 5.3). The three core sequences stay together for the whole range of p-values up to the error threshold that is reached in terms of level crossing at: $p_{1/100}(\mathbf{X}_{79}) = 0.007649$, $p_{1/100}(\mathbf{X}_{207}) = 0.007462$, and $p_{1/100}(\mathbf{X}_{143}) = 0.007704$. It is not easy to guess that four out of the five nearest neighbor sequences have identical values in the eigenvector of the adjacency matrix – and it is the tail-free sequence, $\mathbf{X}_{135}$, and not the sequence carrying the tail, $\mathbf{X}_{175}$, which is different from the other four.

Further increase in the degree of neutrality, λ, gives rise to extended neutral networks, which eventually percolate whole sequence space. Whether or not such large clusters of neutral sequences play a role in real biology cannot be said in the moment but more empirical knowledge on fitness landscapes will help to decide this question.

6. Molecules *in vitro* and prokaryotes

6.1 Molecules in cell-free replication assays

6.2 Viroids and viruses

6.3 Bacteria under controlled conditions

7. Evolution of eukaryotes

8. Probability and stochasticity

An experimentalist reproduces an experiment. What can he expect to find? There are certainly limits to precision and these limits confine the reproducibility of experiments and at the same time restrict the predictability of outcomes. The limitations of correct predictions are commonplace: We witness them every day by watching the failures of various forecasts from the weather to the stock market. Daily experience also tells us that there is an enormous variability in the sensitivity of events with respect to precision of observation. It ranges from the highly sensitive and hard to predict phenomena like the ones just mentioned to the enormous accuracy of astronomical predictions, for example, the precise dating of the eclipse of the sun in Europe on August 11, 1999. Most cases lie between these two extremes and careful analysis of unavoidable randomness becomes important. In this chapter we are heading for a formalism, which allows for proper accounting of the limitations of the deterministic approach and which extends the conventional description by differential equations.

Modeling by differential equations makes one implicit assumption: The use of continuous variables, $x_j(t)$, implies very large, in principle infinitely large populations. Particles are discrete and countable objects and taking into account the non-continuous nature of variables would be quite natural therefore. The reason for using ODEs and PDEs instead is threefold: (i) Large population sizes of 10^{20} particles and more do not allow for counting particles because no experimental measuring technique exist, which could possibly reach the required resolution, (ii) by equilibrium thermodynamics stochastic fluctuation in populations of N particles are of the order $\sqrt{N}$ and $\sqrt{10^{20}} = 10^{10}$ particles is hardly detectable in populations of 10^{20} and more, and (iii) last but not least the experience of modeling by means of differential equations is enormous compared to other mathematical techniques – Leibniz and Newton invented calculus almost 450 years ago.

In biology we can not always assume large particle numbers although populations may reach appreciable sizes, for example the current human population with 7×10^9 individuals. In contrary, numbers are often very small inevitably: Every mutation after all starts out from a single copy, regulatory molecules are often present in a few copies only and due to recombination all individuals in a population have different genomes with the only exception of identical twins – this fact is used for unambiguous identification of individuals in forensic genomics – to give just a few examples. Three different kinds of models with discrete variables are important in biology: (i) Discrete time and continuous population variables for modeling discrete generations obtained through synchronization by day and night rhythms, by seasons or by other pace makers lead to difference equations instead of differential equations, (ii) continuous time and discrete population variables are typically used for populations with small particle numbers and mixing generations and modeled by a variety of stochastic processes, (iii) discrete time and discrete particle numbers. Modeling deterministic discrete time processes is commonly done by means of difference equations [44]. Modeling according to (ii) and (iii) clearly requires probabilistic methods. For answering questions concerning the three examples mentioned above and other similar cases of interest in evolution research the modeling approach (ii) is most appropriate and we shall in essence restrict here the discussion to continuous time stochastic processes.

The conventional description by time dependent variables and functions has to be replaced by a probabilistic view. We are no longer dealing with exact values of variable but we shall use probabilities that variables take on certain values. For example, the number of particles $\mathbf{X}$ is given by the time dependent *stochastic* or *random variable* $\mathcal{X}(t)$,[1] and this implies that the values of $\mathcal{X}$ measured at some time, $t = t_r = t_0 + \Delta_r t$ with t_0 being the initial time, will be different for different repetitions of an experiment under identical conditions. In modern probability theory a rigorous definition of

[1]The notion of a random variable centers on the fact that the value of the variable depends on the outcome of a random process. If an experiment is repeated under identical conditions the value of a random variable will be different for every repetition.

random variables is given on the basis of sets on sample space.

Two introductory sections present the basic notations and definitions of probability theory and stochastic processes for discrete and continuous random variables. Then master equations as used in modeling chemical kinetics will be presented. The advantage of modeling by means of chemical master equations is twofold: (i) Since chemical master equations are based on chemical kinetics they are very general and can be used for nonlinear and complex reaction mechanisms, and (ii) numerical methods are available for the exact simulation of the processes described by chemical master equations [117]. We shall also make a digression into the field of birth-and-death processes, because this class of stochastic processes is most relevant for modeling evolution. Eventually, we introduce and discuss Motoo Kimura's neutral theory of evolution.

8.1 Probabilities and probability theory

The concept of probability originated from the desire to analyze gambling by rigorous mathematical thoughts. An early study that has largely remained unnoticed but contained already the basic ideas of probability was done in the sixteenth century by the Italian mathematician Gerolamo Cardano. The beginning of classical probability theory – 100 years after Cardano – is associated with the story of French mathematician Blaise Pascal and the professional gambler, the Chevalier de Méré that is told in almost ever introduction to probability theory (see section 8.1.1). Classical probability theory can handle all cases that are modeled by discrete quantities. It is based on counting and accordingly fails when applied to uncountable sets. Uncountable sets are occur with continuous variables and these are indispensable for migration in space as well as for large particle numbers that are described in terms of concentrations. Current probability theory is based on set theory and can handle discrete (and countable) as well as continuous (and uncountable) variables. We illustrate the historical probability theory my means of a few examples and then provide a short introduction to modern probabilities that are derived from set theoretical operations. In the third part we present

probability distributions and some of their properties that will be required in the applications to stochastic processes.

8.1.1 Historical probability

The route to probability theory originated from risk and chance estimates in gambling. The oldest well preserved document on a gambling problem making correct use of probability theory is a letter of July 29, 1654, written by Blaise Pascal to Pierre de Fermat in which he reports the careful observation of a professional gambler, the Chevalier de Méré, which let . The Chevalier observed that obtaining at least one "six" with one die in 4 throws is successful in more than 50% in the cases whereas obtaining at least two times the "six" with two dice in 24 throws has less than 50% chance to win. Chevalier de Méré considered this finding as a paradox because he 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 the probabilities correctly as we do it now in classical probability theory by careful counting of events:

$$\text{probability} \;=\; \text{Prob} \;=\; \frac{\text{number of favorable events}}{\text{total number of events}} . \tag{8.1}$$

A probability according to Equ. (8.1) is always a positive quantity between zero and one, $0 \leq \text{Prob} \leq 1$. The sum of the probabilities that an event has occurred or did not occur thus has to be always one. Sometimes, as in the gambler's example, it is easier to calculate the probability of the unfavorable case, q, and to obtain the desired probability as $p = 1 - q$. In the one-die example the probability not to throw a "six" is 5/6, in the two-dice example we have 35/36 as the probability of failure. In case of independent events probabilities are multiplied[2] and we finally obtain for 4 and 24 trials,

[2]We shall come back to the problem of independent events later when we introduce current probability theory in subsection 8.1.2, which is based on set theory.

respectively:

$$q(1) = \left(\frac{5}{6}\right)^4 \quad \text{and} \quad p(1) = 1 - \left(\frac{5}{6}\right)^4 = 0.5177\ ,$$
$$q(2) = \left(\frac{35}{36}\right)^{24} \quad \text{and} \quad p(2) = 1 - \left(\frac{35}{36}\right)^{24} = 0.4914\ .$$

It is remarkable that the gambler could observe this rather small difference in the probability of success – he must have tried the game very often indeed!

Classical probability theory centers around notions like independence of events, joint and conditional probabilities, and many others, which are also central issues of modern set theory based probability and we dispense here from unnecessary repetition.

8.1.2 Sets, sample spaces, and probability

Sets are collections of objects with two restrictions: (i) Each object belongs to one set cannot be a member of 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 application to probability theory we shall denote the elementary objects by the small Greek letter *omega*, ω – if necessary with various sub- and superscripts – and call them *sample points* or *individual results* and the collection of all objects ω under consideration the *sample space* is denoted by Ω with $\omega \in \Omega$. *Events*, A, are subsets of sample points that fulfil some condition[3]

$$A = \left\{\boldsymbol{\omega}, \omega_k \in \Omega : \ f(\boldsymbol{\omega}) = c\right\} \tag{8.2}$$

with $\boldsymbol{\omega} = \{\omega_1, \omega_2, \ldots\}$ being some set of individual results and $f(\boldsymbol{\omega}) = c$ encapsulates the condition on ensemble of the sample points ω_k.[4]

[3]What a *condition* means will become clear later. For the moment it is sufficient to understand a condition as a function providing a restriction, which implies that not all subsets of sample points belong to A.

[4]In order to avoid misinterpretations we shall use different symbols for sets and strings: $\{\cdots\}$ is a set, every element appears only once and the sequence of elements is irrelevant, and $(\alpha_2, \alpha_1, \alpha_1, \cdots)$ is a string where repetitions may occur and the sequence matters. A nucleotide sequence $(\mathbf{GAA}\cdots)$ is a string and the nucleotide alphabet is a (sub)set: $\mathcal{A} = \{\mathbf{A}, \mathbf{U}, \mathbf{G}, \mathbf{C}\}$.

Any partial collection of points 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 some set S is called its size, $|S|$, and thus is a nonnegative integer or ∞. In particular, the size of the empty set is $|\emptyset| = 0$. The unambiguous assignment of points to sets can be expressed by[5]

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

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:

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

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

The basic operations with sets are illustrated in Fig. 8.1. 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:[6]

$$A^c \ = \ \{\omega | \omega \notin A\} \ . \tag{8.3}$$

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:

$$A \cup B \ = \ \{\omega | \omega \in A \text{ or } \omega \in B\} \ . \tag{8.4}$$

Intersection. The intersection of the two sets A and B, $A \cap B$, is the set of points which belong to both sets (For short $A \cap B$ is sometimes written AB):

$$A \cap B \ = \ \{\omega | \omega \in A \text{ and } \omega \in B\} \ . \tag{8.5}$$

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

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

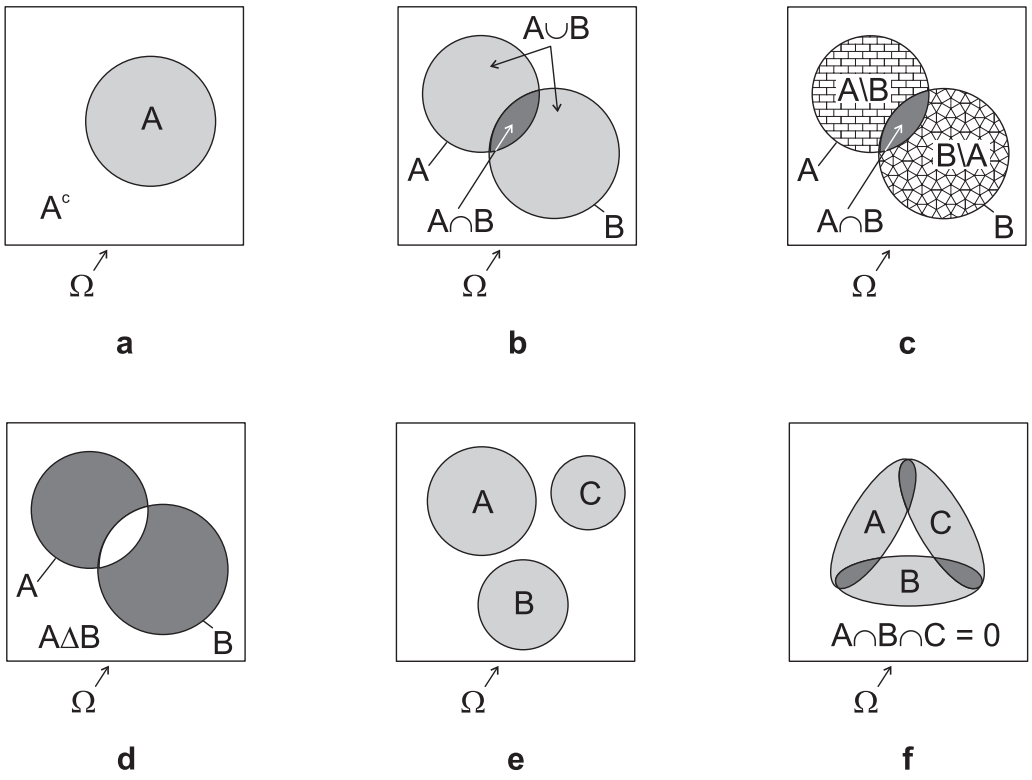


Figure 8.1: **Some definitions and examples from set theory.** Part **a** shows the complement A^c of a set A in the sample space Ω. In part **b** we explain the two basic operations union and intersection, $A \cup B$ and $A \cap B$, respectively. Parts **c** and **d** show the set-theoretic difference, $A \setminus B$ 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.

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

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

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

These relations are true 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 $A \setminus B$ is the set of points, which belong to A but not to B:

$$A \setminus B = A \cap B^c = \{\omega | \omega \in A \text{ and } \omega \notin B\} \, . \tag{8.6}$$

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, $A\Delta B = B\Delta A$:

$$A\Delta B = (A \cap B^c) \cup (A^c \cap B) = (A \setminus B) \cup (B \setminus A) \, . \tag{8.7}$$

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:

$$A \cap B = \emptyset \, , \quad A \subset B^c \text{ and } B \subset A^c \, . \tag{8.8}$$

A number of 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 *decomposition*: $\Omega = A + A^c$.

Sample spaces may contain finite or infinite numbers of sample points. As shown in Fig. 8.2 it is important to distinguish further between different classes of infinity: *countable* and *uncountable* numbers of points. The set of rational numbers, for example, is a countably infinite since the numbers can be labeled and assigned uniquely to the positive integers

$$1 < 2 < 3 < \cdots < n < \cdots .$$

The set of real numbers cannot be ordered in such a way and hence it is uncountable.

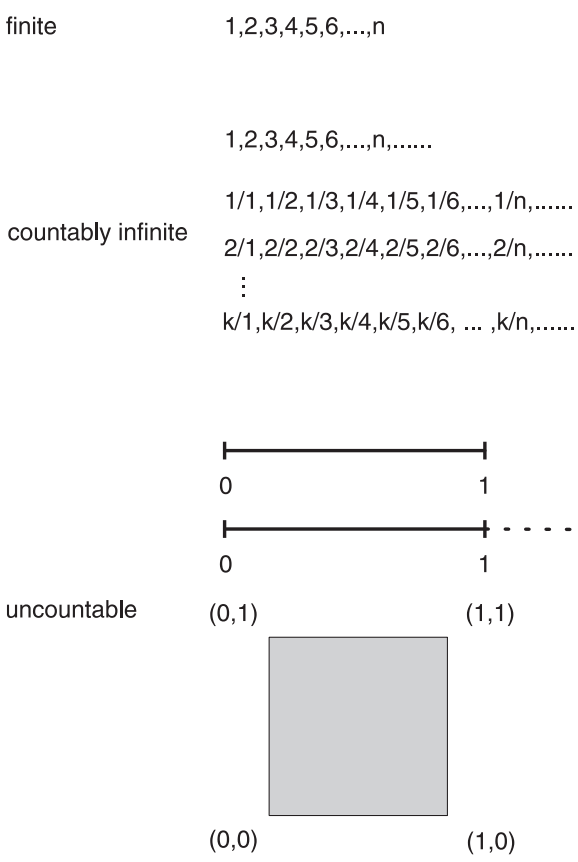


Figure 8.2: **Sizes of sample sets and countability.** Finite, countably infinite, and uncountable sets are distinguished. We show examples of every class. A set is countably infinite when its elements can be assigned uniquely to the natural numbers (1,2,3,...,n,...).

Countable sample spaces. For countable sets it is straightforward to measure the size of sets by counting the numbers of points they contain. The proportion

$$\text{Prob}\,(A) \;=\; P(A) \;=\; \frac{|A|}{|\Omega|} \tag{8.9}$$

is identified as the probability of the event represented by the elements of subset A.[7] For countable and finite sets this is precisely Equ. (8.1). For another event holds, for example, $P(B) = |B|/|\Omega|$. Calculating the sum of

[7]The number of elements or the *cardinality* of a set A is denoted by $n(A)$ or $|A|$.

the two probabilities, $P(A) + P(B)$, requires some care since we know only (Fig. 8.1):

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

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

$$P(A) \; + \; P(B) \; = \; P(A \cup B) \; + \; P(A \cap B) \; .$$

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

$$P(A + B) \; = \; P(A) \; + \; P(B) \;\; \text{iff} \;\; A \cap B = \emptyset \; . \tag{8.10}$$

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

Now we can define a probability measure by means of the basic axioms of probability theory (for alternative axioms in probability theory see, for example [40, 164]):

A *probability measure* on the sample space Ω is a function of subsets of Ω, $P : S \to P(S)$ or $P(\cdot)$ for short, 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 value for A and its value for B,

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

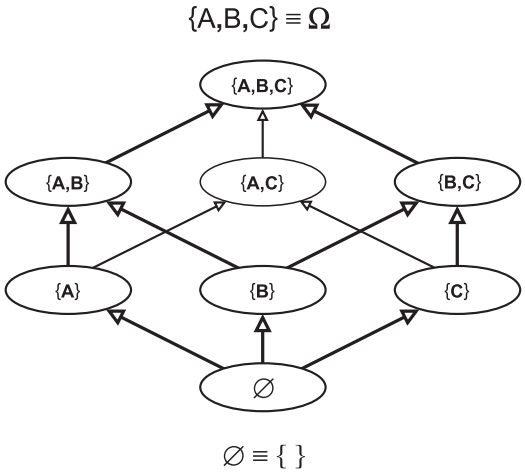


Figure 8.3: **The powerset.** The powerset $\mathcal{P}(\Omega)$ is a set containing all subsets of Ω including the empty set $\emptyset$ and Ω itself. The figure sketches the powerset of three events {A, B, C}.

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 $i \neq j$, the relation called *σ-additivity* or *countable additivity*

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i) \quad \text{or} \quad P\left(\sum_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) \tag{8.11}$$

holds. 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 have $P(A) \leq P(B)$ and $P(B - A) = P(B) - P(A)$. For any two arbitrary sets A and B we can write a sum of disjoint sets as follows

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

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 the *powerset* $\mathcal{P}(\Omega)$ (Fig. 8.3). It contains the empty set $\emptyset$, the sample space Ω and all subsets of Ω and this includes the results of all set theoretic operations that were listed above. The powerset

of some given set S has cardinality $2^{|S|}$. The relation between the sample point ω, an event A, the sample space Ω and the powerset $\mathcal{P}(\Omega)$ is illustrated by means of an example taken from molecular evolution of nucleic acid sequences: single nucleotide exchanges or point mutations. Mutation in natural DNA or RNA sequences may be considered like rolling dice with four outcomes, e.g. $\mathbf{A}\rightarrow\mathbf{T}(\mathbf{U})$, $\mathbf{A}\rightarrow\mathbf{G}$, and $\mathbf{A}\rightarrow\mathbf{C}$, and the fourth outcome is $\mathbf{A}\rightarrow\mathbf{A}$, no change or correct reproduction. For the sake of simplicity we assume binary sequences that correspond to Bernoulli trials discussed in subsection 8.2.1. Events ω are mutations $\mathbf{0}\rightarrow\mathbf{1}$, and the sample points for flipping the coin n-times are binary n-tuples or strings, $\omega=(\omega_1,\omega_2,\ldots,\omega_n)$ with $\omega_i\in\{\mathbf{0},\mathbf{1}\}$.[8] The powerset shown in Fig. 8.3 corresponds to sequences of length $l=3$ with the empty set, $\emptyset$, is represented by (**000**). Events indicate the appearance of a **1** at a certain position: $A\equiv\mathbf{100}$, $B\equiv\mathbf{010}$, $C\equiv\mathbf{001}$, $AB\equiv\mathbf{110}$, etc. Identifying some (reference) binary sequence $\mathbf{X}_0$ with the empty set the analogy between the power set and the sequence space $\mathcal{Q}$ is recognized straightforwardly. It is useful to consider also infinite numbers of repeats, in particular for computing limits $n\rightarrow\infty$: $\omega=(\omega_1,\omega_2,\ldots)=(\omega_i)_{i\in\mathbb{N}}$ with $\omega_i\in\{0,1\}$. Then we are dealing with infinitely long binary strings and the sample space $\Omega=\{0,1\}^{\mathbb{N}}$ is the space of all infinitely long binary strings. It is countable as can be easily verified: Every binary string represents the binary encoding of a natural number (including '0') $N_k\in\mathbb{N}_0$ and hence Ω is countable as the natural numbers are.

A subset of Ω will be called an event A iff 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 coarse grained by lumping together sample points. We ask, for example, for the probability A that n coin flips yield at least k-times *tail*, i.e. the score 1:

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

[8] There is a trivial but important distinction between strings or n-tuples 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 string are written in (normal) parentheses and sets in *curly brackets*.

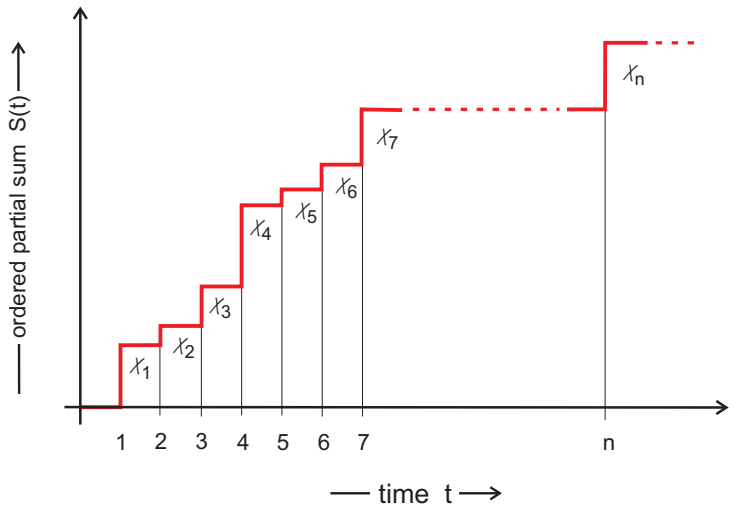


Figure 8.4: **An ordered partial sum of a random variable.** The partial sum $S_n = \sum_{k=1}^{n} \mathcal{X}_k$ represents the cumulative outcome of a series of events described by a class of random variables, $\mathcal{X}_k$. The series can be extended to $+\infty$ and such a case will be encountered, for example, with probability distributions, and when the series is completed the sum has to fulfil: $S_n = \sum_{k=1}^{\infty} \mathcal{X}_k = 1$. The ordering criterion is not yet specified, it could be time t, for example.

where the sample space is $\Omega = \{0, 1\}^n$. The task is now to find a system of events $\mathcal{F}$ that allows for a consistent assignment of a probability $P(A)$ for every event A. For countable sample spaces Ω the powerset $\mathcal{P}(\Omega)$ represents such a system $\mathcal{F}$, we characterize $P(A)$ as a probability measure on $\big(\Omega, \mathcal{P}(\Omega)\big)$, and the further handling of probabilities as outlined below is straightforward. In case of uncountable sample spaces Ω, however, the powerset $\mathcal{P}(\Omega)$ is too large and a more sophisticated procedure is required (see, e.g., Schuster:11d). Weighting sample points ω_n by assigning factors ϱ_n eventually leads to a probability functions, which assign local probabilities or cumulative probabilities to events and which are called (probability) densities and distributions, respectively (see subsection 8.1.3).

Random variables and functions. For a precise definition of random variables on countable sets a *probability triple* $(\Omega, \mathcal{P}(\Omega), P)$ is required: Ω contains the sample points or individual results, the powerset $\mathcal{P}(\Omega)$ provides the events A as subsets, and P eventually represents a probability measure that will be precisely defined in Equ. (8.19). Based on such a probability triple we define a *random variable* as a numerically valued function $\mathcal{X}$ of ω on the domain of the entire sample space Ω,

$$\omega \in \Omega : \quad \omega \rightarrow \mathcal{X}(\omega) \; . \tag{8.12}$$

Random variables, $\mathcal{X}(\omega)$ and $\mathcal{Y}(\omega)$, can be subject to 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, also any linear combination of random variables such as $\alpha\mathcal{X}(\omega) + \beta\mathcal{Y}(\omega)$ is a random variable too. Just as a function of a function is still a function, a function of a random variable is a random variable,

$$\omega \in \Omega : \quad \omega \rightarrow \varphi\left(\mathcal{X}(\omega), \mathcal{Y}(\omega)\right) = \varphi(\mathcal{X}, \mathcal{Y}) \; .$$

Particularly important cases are the (partial) sums of n variables:

$$S_n(\omega) \; = \; \mathcal{X}_1(\omega) \, + \, \ldots \, + \, \mathcal{X}_n(\omega) \; = \; \sum_{k=1}^{n} \mathcal{X}_k(\omega) \; . \tag{8.13}$$

Such a partial sum S_n could be, for example, the cumulative outcome of n successive throws of a die.[9] Consider, for example, an ordered series of events where the current cumulative outcome is given by the partial sum $S_n = \sum_{k=1}^{n} \mathcal{X}_k$ as shown in Fig. 8.4. In principle, the series can be extended to infinity and then the conservation relation of probabilities, $S_n = \sum_{k=1}^{\infty} \mathcal{X}_k = 1$, has to be fulfilled.

[9]The use of *partial* in this context implies that the sum does not cover the entire sample space in the moment. Series of throws of dice, for example, could be continued in the future.

Joint probabilities. Two (or more) random variables,[10] $\mathcal{X}$ and $\mathcal{Y}$, form a *random vector*, $\vec{\mathcal{Z}} = (\mathcal{X}, \mathcal{Y})$, which is defined by the probability

$$P(\mathcal{X} = x_i, \mathcal{Y} = y_j) \;=\; p(x_i, y_j) \;. \tag{8.14}$$

These probabilities constitute the *joint probability distribution* (see subsection 8.1.3) of the random vector. By summation over one variable we obtain the probabilities for the two *marginal distributions*:

$$\begin{aligned} P(\mathcal{X} = x_i) \;&=\; \sum_{y_j} p(x_i, y_j) \;=\; p(x_i, *) \;\text{ and} \\ P(\mathcal{Y} = y_j) \;&=\; \sum_{x_i} p(x_i, y_j) \;=\; p(*, y_j) \;, \end{aligned} \tag{8.15}$$

of $\mathcal{X}$ and $\mathcal{Y}$, respectively. In other words, the marginal distributions reduce the multivariate function $p(x_i, y_j)$ to functions of single variables, $p(x_i)$ and $p(y_j)$ through summation over all possible values of the second variable.

Conditional probabilities. The conventional probability is defined on the entire sample space Ω, $P(A) = |A|/|\Omega| = \sum_{\omega\in A} P(\omega) \,\big/\, \sum_{\omega\in\Omega} P(\omega)$.[11] We shall now define a probability of set A relative to another set, say S. This means that we are interested in the proportional weight of the part of A in S which is expressed by the intersection $A \cap S$ relative to 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 consider the *conditional probability of A relative to S*:

$$P(A|S) \;=\; \frac{P(A\cap S)}{P(S)} \;=\; \frac{P(AS)}{P(S)} \tag{8.16}$$

provided $P(S) \neq 0$. From here on we shall always use the a short notation for the intersection, $AS \equiv A \cap S$, in this section. Apparently, the conditional probability vanishes if the intersection is empty: $P(A|S) = 0$ if $P(AS) = \emptyset$.

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

[11] The sample space Ω is assumed to be countable and the weight $P(\omega) = P(\{\omega\})$ is assigned to every point.

Next we mention several simple but fundamental relations involving conditional probabilities that we present here, in essence, without proof (for details see [36], pp.111-144). For n arbitrary events A_i we have

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

provided $P(A_1A_2\ldots A_{n-1}) > 0$. Under this proviso all conditional probabilities are well defined since

$$P(A_1) \geq P(A_1A_2) \geq \ldots \geq P(A_1A_2\ldots A_{n-1}) > 0\ .$$

Let us assume that the sample space Ω is partitioned into n disjoint sets, $\Omega = \sum_n A_n$. For any set B we have then

$$P(B) \;=\; \sum_n P(A_n)\,P(B|A_n)\ .$$

From this relation it is straightforward to derive the conditional probability

$$P(A_j|B) \;=\; \frac{P(A_j)P(B|A_j)}{\sum_n P(A_n)P(B|A_n)}$$

provided $P(B) > 0$.

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:

$$P(\mathcal{X}_1 = x_1, \ldots, \mathcal{X}_n = x_n) \;=\; P(\mathcal{X}_1 = x_1) \cdot \ldots \cdot P(\mathcal{X}_n = x_n)\ . \qquad (8.17)$$

A major extension of Equ. (8.17) 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

$S_1, \ldots, S_n$:

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

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

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

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

8.1.3 Probability distributions

Depending on the nature of variables probability distributions are discrete or continuous. We introduce probability distributions be discussing the conceptually simpler discrete case and present then the differences encountered for continuous probability functions.

Discrete probabilities. So far we have constructed and compared sets but not yet introduced numbers, which are needed in actual computations. In

order to construct a probability measure that is adaptable for numerical calculations on countable sample spaces, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$, we assign a weight ϱ_n to every sample point ω_n subject to the conditions

$$\forall\ n: \quad \varrho_n \geq 0\,;\ \sum_n \varrho_n = 1\ . \tag{8.18}$$

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

$$\begin{aligned} P(A) &= \sum_{\omega\in A} \varrho(\omega) \quad \text{for} \quad A \in \mathcal{P}(\Omega) \quad \text{and} \\ \varrho(\omega) &= P(\{\omega\}) \quad \text{for} \quad \omega \in \Omega \end{aligned} \tag{8.19}$$

represent a bijective relation between the probability measure P on $\big(\Omega, \mathcal{P}(\Omega)\big)$ 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 *probability density* or *probability mass function*.

The function $\varrho(\omega_n) = \varrho_n$ has to be estimated or determined empirically because it is the result of factors lying outside mathematics or probability theory. In physics and chemistry the correct assignment of probabilities has to meet the conditions of the experimental setup. An example will make this point clear: The fact whether a die is fair and shows all its six faces with equal probability or it has been manipulated and shows the "six" more frequently then the other numbers is a matter of physics and not mathematics. For many purposes the discrete *uniform distribution*, Π_Ω, is applied: All results $\omega \in \Omega$ appear with equal probability and hence $\varrho(\omega) = 1/|\Omega|$. In Fig. 8.5 the scores obtained by simultaneously rolling two dice are shown: The events are independent and hance the two scores based on the assumption of uniform probabilities are added.

The probabilistic nature of random variables is illustrated well by a different formulation, which is particularly useful for the definition of probability distribution functions:

$$P_k(t) = \text{Prob}\big(\mathcal{X}(t) = k\big) \quad \text{with} \quad k \in \mathbb{N}^0\ . \tag{8.20}$$

In a stochastic process we can properly visualize the change in the representation of the process as the replacement of a deterministic variable $x(t)$ by an

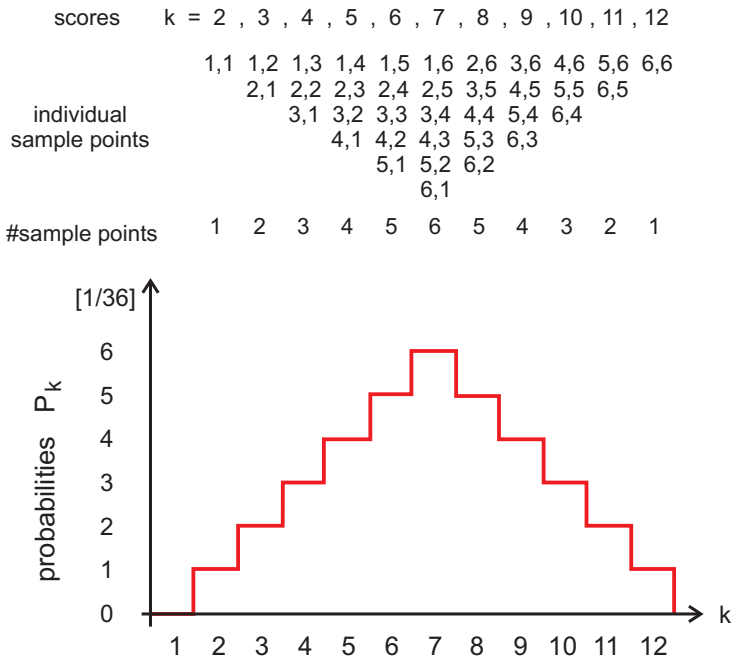


Figure 8.5: **Probabilities of throwing two dice.** The probability of obtaining scores from two to twelve counts through throwing two fair dice are based on the equal probability assumption (uniform distribution Π) for obtaining the individual faces of a single die. The probability mass function $f(k) = P_k$ raises linearly from two to seven and then decreases linearly between seven and twelve ($P(N)$ is a discretized tent map) and the additivity condition requires $\sum_{k=2}^{12} P(k) = 1$.

evolving probability vector $\boldsymbol{P}(t) = \big(P_0(t), P_1(t), \ldots\big)$.[12] It is worth noticing that two separate changes are introduces here simultaneously: (i) the continuous concentration is replaced by a discrete particle number, and (ii) the deterministic description is substituted by a probabilistic view. The elements P_k are probabilities hence they must fulfil two conditions: (i) they have to be nonnegative numbers, $P_k \geq 0$, and (ii) the set of all possible events covers the sample space Ω and fulfils the conservation relation, which can be formulated

[12] Whenever possible we shall use "k, l, m, n" for discrete counts, $k \in \mathbb{N}^0$, and "x, y, z" for continuous variables, $x \in \mathbb{R}^1$.

also in terms of classical probability theory

$$P_k \;=\; \text{Prob}\,\frac{\text{favorable events}}{\text{all events}} \;\text{ leading to }\; \sum_{i=1}^{n} P_i \;=\; 1\;.$$

Recalling the mutation-selection equation (4.9) we may notice now that the relation $\sum_{i=1}^{n} Q_{ij} = 1$ can be interpreted as conservation of probabilities.

Discrete probabilities for complete sets of events are represented either by the *probability mass function* (pmf)

$$f_{\mathcal{X}}(k) \;=\; \text{Prob}\,(\mathcal{X} \;=\; k) \;=\; p_k \tag{8.21}$$

or by the *cumulative distribution function* (cdf)

$$F_{\mathcal{X}}(k) \;=\; \text{Prob}\,(\mathcal{X} \;\leq\; k) \;=\; \sum_{i\leq k} p_i \tag{8.22}$$

Two properties of the cumulative distribution function are self evident:

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

The limit at low k-values is chosen in analogy to the definitions: Taking zero instead of $-\infty$ as lower limit because $f_{\mathcal{X}}(-|k|) = p_{-|k|} = 0$ ($k \in \mathbb{N}$) or, in other words, negative particle numbers have zero probability (For more details on probability functions see [36, 251]). A simple example of probability functions is shown in Fig. 8.7. All measurable quantities can be computed from either of the two probability functions.

As seen in Fig. 8.7 the cumulative distribution function is a step function and requires a convention how the steps are treated in detail. Indeed, step functions have limited or no differentiability depending on the definition of the values of the function at the step. Three definitions are possible for the value of the function at the discontinuity. We present them for the Heaviside step function

$$H(k) \;=\; \begin{cases} 0\,, & \text{if } k < 0\,, \\ 0,\, \frac{1}{2},\, 1 & \text{if } k = 0\,, \\ 1\,, & \text{if } k > 0\,. \end{cases} \tag{8.23}$$

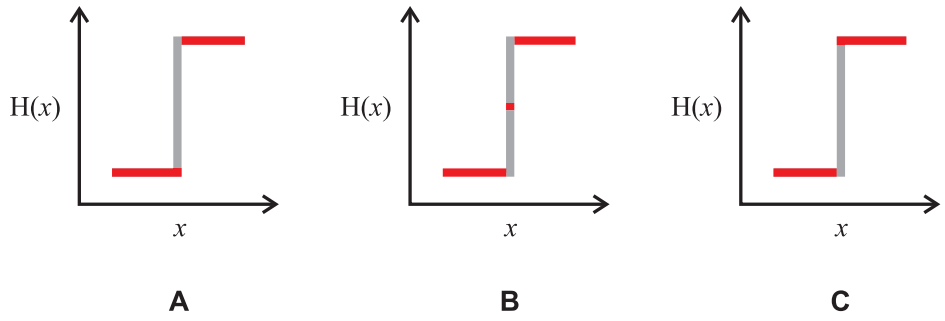


Figure 8.6: **Continuity in probability theory and step processes.** Three possible choices of partial continuity at the steps of step functions are shown: (i) left-hand continuity (**A**), (ii) no continuity (**B**), and (iii) right-hand continuity (**C**). The step function in (**A**) is left-hand 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) allows for making use of the inherent symmetry of the Heaviside function. Choice (iii) is the standard assumption in probability theory and stochastic processes. It is also known as *càdlàg*-property (subsection 8.2.1).

The value '0' at $k = 0$ implies left-hand continuity for $H(k)$ and in terms of a probability distribution would correspond to a definition $\text{Prob}\,(\mathcal{X} < k)$ in Equ. (8.22), the value $\frac{1}{2}$ implies that $H(k)$ is neither right-hand nor left-hand semi-differentiable at $k = 0$ but is useful in many applications that make use of the inherent symmetry of the Heaviside function, for example the relation $H(x) = \big(1 + \text{sgn}(x)\big)/2$ where $\text{sgn}(x)$ is the sign or signum function:

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

The functions in probability theory make use of the third definition determined by $P(\mathcal{X} \leq x)$ or $H(0) = 1$ in case of the Heaviside function. This choice implies right-hand continuity or right-hand semi-differentiability and is important in the conventional handling of stochastic processes.

Fig. (8.7) presents the probability mass function (pmf) and the cumulative probability distribution function (cdf) for the scores of rolling two dice. The

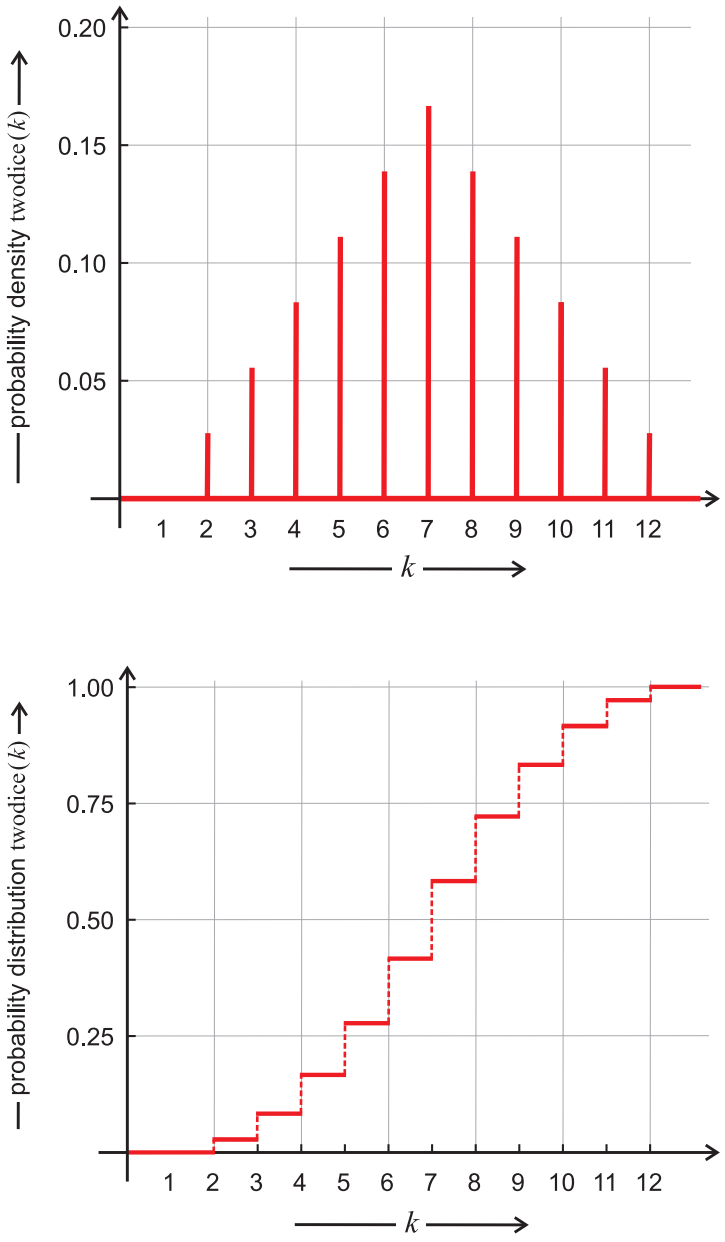


Figure 8.7: **Probability mass function and cumulative distribution function for rolling two dice.** As as example we show here the probabilities for the scores obtained by rolling two dice simultaneously with the complete set of events $\Omega = \{2, 3, \ldots, 12\}$ and the random variable $\mathcal{X} \in \Omega$. The probability mass function (pmf; upper plot) has its maximum at the score "7", because it is obtained with the maximum number of combinations (1+6, 2+5, 3+4, 4+3, 5+2, 6+1) leading to a probability of $6/36 = 1/6 = 0.1667$. The cumulative distribution function (cdf; lower plot) presents the summation of all probabilities of lower scores: $F_{\mathcal{X}}(k) = P(\mathcal{X} \leq k) = \sum_{i \leq k} p_i$ and $\lim_{k \to +\infty} F_{\mathcal{X}}(k) = 1$.

former is given by the tent function

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

where 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 is given by the sum

$$F(k) \;=\; \sum_{i=2}^{2s} f(i)\;.$$

Here the two-dice probability distribution is used only as an example for the illustration of probability functions, but later on we shall extend to the n-dice score problem in order to illustrate the law of large numbers.

The Poisson distribution. The Poisson distribution, a discrete probability distribution, is of particular importance in the theory of stochastic processes, because it expresses the probability that a given number of events occurs within a fixed time interval (The corresponding stochastic process is the Poisson process discussed in subsection 9.1.2). The events take place with a known average rate and independently of the time that has elapsed since the last event. An illustrative example is the arrival of (independent) e-mails: We assume a person receiving on the average 50 e-mails per day, then the Poisson probability mass function

$$\text{Pois}(k;\lambda): \quad f(k) \;=\; \frac{\lambda^k}{k!}\,e^{-\lambda} \;\text{ with }\; k \;\in\; \mathbb{N}^0 = \{0,1,2,\ldots\}\;, \tag{8.24}$$

returns the probability that exactly x are arriving in the same time interval, i.e. per day. The corresponding cumulative distribution function is

$$\text{Pois}(k;\lambda): \quad F(k) \;=\; \frac{\Gamma(\lfloor k+1\rfloor,\lambda)}{\lfloor k\rfloor!} \;=\; e^{-\lambda}\sum_{k=0}^{\lfloor k\rfloor}\frac{\lambda^k}{k!} \;\text{ for }\; k \;\leq\; 0\;, \tag{8.25}$$

with $\Gamma(s,z)$ being the upper and $\gamma(s,z)$ the lower incomplete Γ-function:

$$\Gamma(s,z) \;=\; \int_z^\infty x^{s-1}\,e^{-x}\,\mathrm{d}x\,,\;\; \gamma(s,z) \;=\; \int_0^z x^{s-1}\,e^{-x}\,\mathrm{d}x\,,\;\; \Gamma(s) = \Gamma(s,z) + \gamma(s,z)\;,$$

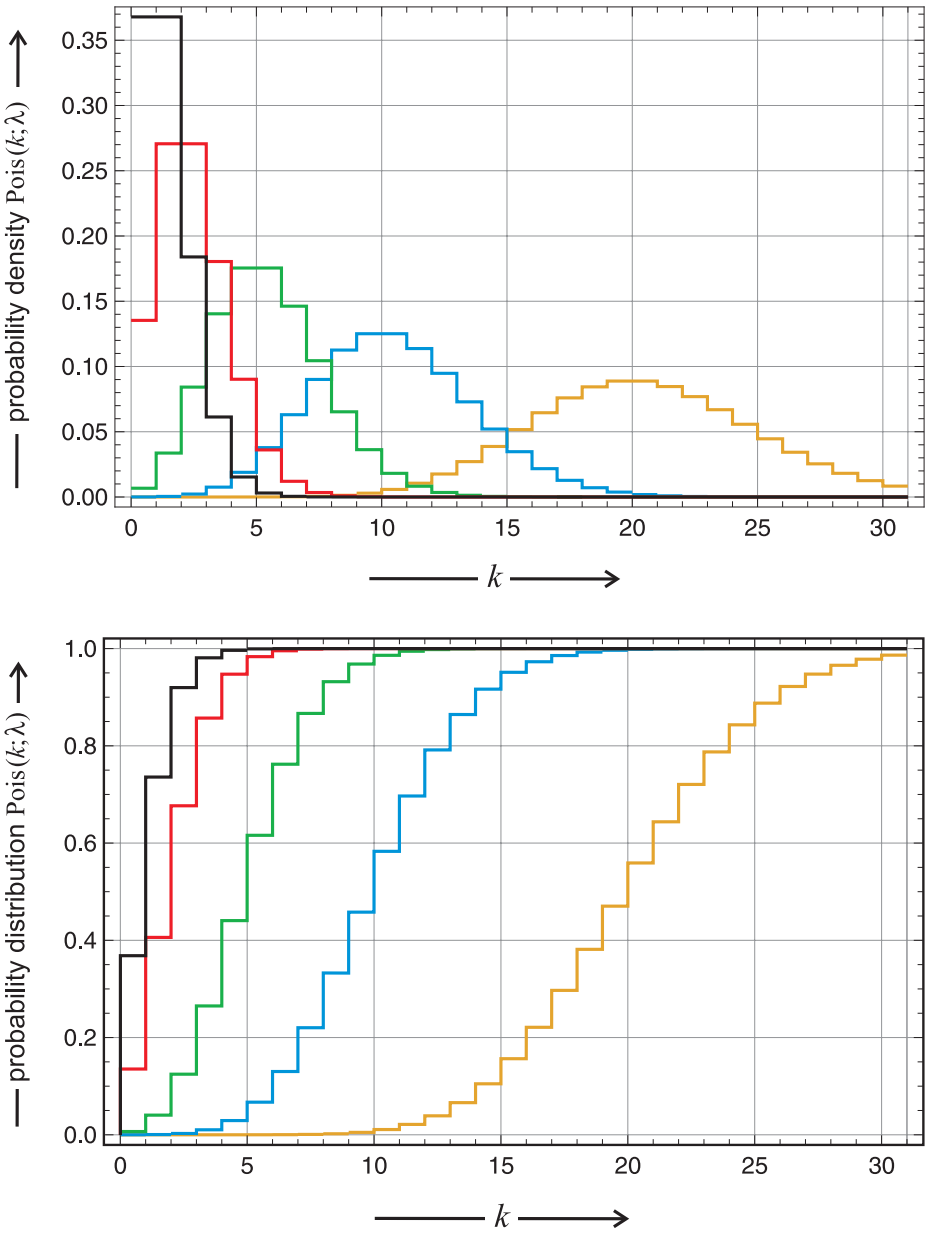


Figure 8.8: **Poisson density and distribution.** In the plots the Poisson distribution, Pois(λ), is shown in from of the probability density $f(k)$ and the probability distribution $F(k)$ as an example. The probability density $f(k)$ is asymmetric for small λ-values ($\lambda = 1, 2, 5$) and steeper on the (left-hand) side of k-values blow the maximum, whereas large λ's give rise to more and more symmetric curves as expected by the law of large numbers. Choice of parameters: $\lambda = 1$ (black), 2 (red), 5 (green), 10 (blue) and 20 (yellow).

and $\lfloor k \rfloor$ being the *floor* function, which extracts j being the largest integer i ($i, j \in \mathbb{N}^0$) that fulfils $j = (\max\{i\} \wedge i \leq k)$. Examples of Poisson distributions with different λ-values are shown in Fig. 8.8. Time span is not the only interval for defining the Poisson distribution as the reference for independent events. It can be used equally well for events taking place in space with the intervals referring to distance, area or volume.

In science the Poisson distribution is used for modeling independent events. The most relevant case for applications in chemistry and biology is the occurrence of elementary chemical reactions, which are thought to be initiated by an encounter of two molecules. Since the numbers of molecules are large and the macroscopic dimensions of a reaction vessel exceed the distances between two encounters of a given molecule by many orders of magnitude, reactive collisions can be considered as independent events (see also 9.1.2). Concerning evolutionary aspects a typical biological example to which a Poisson distribution of events applies is the number of mutations per unit time interval on a given stretch of DNA. In subsection 9.3.1 a master equation based on collision theory will be derived for the simulation of chemical reactions.

Eventually, we shall characterize a probability distribution by its *moments* of the random variable $\mathcal{X}$. Most important for practical purposes are the first moment or the mean that takes on the value

$$E(\mathcal{X}) \;=\; \sum_{k=0}^{\infty} k\, p_k \;=\; \sum_{x=0}^{\infty} x\, f(x) \;=\; \lambda \tag{8.26}$$

for the Poisson distribution and the second centered moment or the variance. which becomes

$$\begin{aligned} \sigma^2(\mathcal{X}) \;&=\; E\Big(\big(\mathcal{X} - E(\mathcal{X})\big)^2\Big) \;=\; E(\mathcal{X}^2) \;-\; \big(E(\mathcal{X})\big)^2 \;=\; \lambda \\ &\text{with}\;\; E(\mathcal{X}^2) \;=\; \sum_{x=0}^{\infty} x^2\, f(x)\;. \end{aligned} \tag{8.27}$$

In the Poisson distribution the mean is equal to the variance and takes on the value of the parameter, λ. Often the standard deviation σ is used instead of the variance. Higher moments are used sometimes to characterize further

details of probability distributions we mentioned here the third moment expressed as the skewness and the fourth moment called in the kurtosis of the distribution.

Continuous probabilities. A rigorous introduction of continuous distributions of random variables[13] is quite involved and requires an introduction into uncountable sets and measure theory [36, 251]. We shall dispense here from the mathematical part and make use only of applications. A continuous random variable is commonly defined on the real numbers $\mathcal{X} \in \mathbb{R}^1$. Discrete probability mass functions (pmf) and cumulative distribution functions (cdf) have their analogues in continuous probability density functions (pdf) and cumulative probability distribution functions (cdf), and summation is replaced by integration. The probability density is a function f on $\mathbb{R} =]-\infty, +\infty[$, $u \to f(u)$, which satisfies the two conditions:

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

which are the analogues to the positivity condition and conservation relation of probabilities. Now we can rigorously define random variables on general sample spaces: $\mathcal{X}$ is a function on Ω, which is here the set of real numbers $\mathbb{R}$ and whose probabilities are prescribed by means of a density function $f(u)$. For any interval $[a, b]$ the probability is given by

$$\text{Prob}\,(a \leq \mathcal{X} \leq b) \ = \ \int_a^b f(u)\, du \ . \tag{8.28}$$

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

$$F(x) \ = \ P(\mathcal{X} \leq x) \ = \ \int_{-\infty}^{x} f(u)\, du \ .$$

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

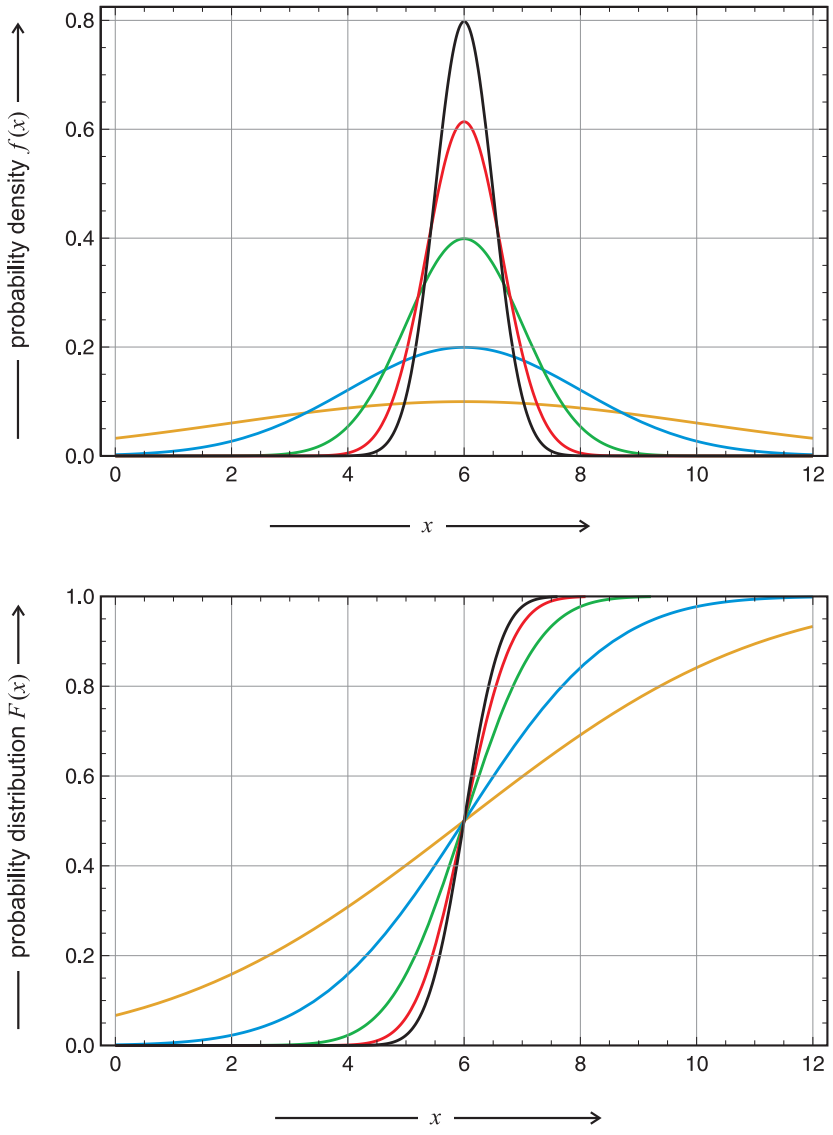


Figure 8.9: **Normal density and distribution.** In the plots the normal distribution, $\mathcal{N}(\mu, \sigma)$, is shown in from of the probability density $f(x) = \exp\Big(-(x-\mu)^2/(2\sigma^2)\Big) \;/\; (\sqrt{2\pi}\,\sigma)$ and the probability distribution $F(x) = \Big(1 + \mathrm{erf}\big((x-\mu)/\sqrt{2\sigma^2}\big) \;/\; 2\Big)$. Choice of parameters: $\mu = 6$ and $\sigma = 0.5$ (black), 0.65 (red), 1 (green), 2 (blue) and 4 (yellow).

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 $a = b = x$

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

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, however, can be resolved by looking at countable and uncountable sets and measures defined on them in more depth [36, 251].

The first moment, the second central moment, and higher central moments are defined for continuous distributions in the same way as for discrete distributions only the summation is replaced by an integral:[14]

$$E(\mathcal{X}) \;=\; \hat{\mu}(\mathcal{X}) \;=\; \int_{-\infty}^{+\infty} x\,f(x)\,\mathrm{d}x\,, \quad \text{and} \tag{8.29a}$$

$$\sigma^2(\mathcal{X}) \;=\; \mu_2(\mathcal{X}) \;=\; \int_{-\infty}^{+\infty} \bigl(x - E(\mathcal{X})\bigr)^2\,f(x)\,\mathrm{d}x \tag{8.29b}$$

$$\mu_n(\mathcal{X}) \;=\; \int_{-\infty}^{+\infty} \bigl(x - E(\mathcal{X})\bigr)^n\,f(x)\,\mathrm{d}x \tag{8.29c}$$

Two higher moments are frequently used to characterize the detailed shape of a distribution, the *skewness* is expressed as the third central moment $\mu_3(\mathcal{X})$ and the *kurtosis* is obtained form the fourth moment $\mu_4(\mathcal{X})$. The skewness describes the asymmetry of the distribution: A density with positive skewness is flatter on the right-hand-side (at larger values of $\mathcal{X}$; for example the Poisson density), whereas one with negative skewness has the steeper slope at the right-hand-side. The kurtosis is a measure of the "peakedness"

[14] *Raw moments* $\hat{\mu}_n(\mathcal{X}) = E(\mathcal{X}^n) = \int_{-\infty}^{+\infty} x^n\,f(x)\,\mathrm{d}x$ are distinguished from *central moments* $\mu_n(\mathcal{X}) = E\Bigl(\bigl(\mathcal{X} - E(\mathcal{X})\bigr)^n\Bigr) = \int_{-\infty}^{+\infty} \bigl(x - \hat{\mu}(\mathcal{X})\bigr)^n\,f(x)\,\mathrm{d}x$, and hence $\mu(\mathcal{X}) = 0$.

of a distribution: More positive kurtosis implies a sharper peak whereas negative kurtosis indicates a flatter maximum of the distribution. Often the attributes "leptokurtic" and "platykurtic" are given to densities with positive and negative excess kurtosis, respectively, where excess refers to values with respect to the normal distribution.

The number of special continuous probability distributions reflecting different circumstances and conditions is, of course, large and we shall mention only two of them, which are of particular importance for probability theory and stochastic processes in general and in particular in chemistry and biology, the normal distribution (Fig. 8.9) and the somewhat pathological *Cauchy-Lorentz distribution* (Fig. 8.11).

The normal distribution. The density of the normal distribution is a Gaussian function named after the German mathematician Carl Friedrich Gauß and is often called symmetric bell curve.

$$\mathcal{N}(x;\mu,\sigma^2): \quad f(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\,\sigma^2}} \;, \tag{8.30}$$

$$F(x) \;=\; \frac{1}{2}\left(1+\mathrm{erf}\Big(\frac{x-\mu}{\sqrt{2\sigma^2}}\Big)\right) \;. \tag{8.31}$$

The two parameters of the normal distribution are at the same time the first and the second moment: the mean, $E(\mathcal{X})=\mu$, and the variance, $\sigma^2(\mathcal{X})=\sigma^2$, and all odd central moments are zero because of symmetry. The even central moments take on the general form

$$\mu_{2n}(\mathcal{X}) \;=\; \frac{(2n)!}{2^n\, n!}\left(\sigma^2\right)^n \;,$$

and we find for the kurtosis of the normal distribution $\mu_4(\mathcal{X})=3\sigma^4$. The kurtosis of general distribution is often expressed as *excess kurtosis* or fourth cumulant, $\kappa_4=\mu_4-3\mu_2^2=\mu_4-3(\sigma^2)^2=\mu_4-3\sigma^4$, that measures kurtosis relative to the normal distribution.

The presumably most important property of the normal distribution is encapsulated in the *central limit theorem*, which says in a nutshell: Every probability distribution converges to the normal distribution provided the

number of sampling points approaches infinity. The central limit theorem can be proved exactly (see [36, 251]) but here we shall make use of two instructive examples of discrete probability distributions converging to the normal distribution at large numbers of sample points. Commonly, the convergence of to the normal distribution is illustrated by means of the *binomial distribution*

$$B(k;p,n):\ f(x)\ =\ \binom{n}{k}p^k\,(1-p)^{n-k}\ \text{ with }\ k=\{0,1,2,\ldots\,n\}\,,\ n\in\mathbb{N}^0\,. \quad (8.32)$$

The binomial distribution is the discrete probability distribution of the numbers of successes within a sequence of n independent trials of binary (yes/no) decisions, each of which has a probability p to succeed. Such trials are called *Bernoulli trials* and the individual trial yields a one (success) with probability p and a zero (failure) with probability $q=1-p$. The binomial distribution $B(k;p,1)$ is the Bernoulli distribution. The cumulative distribution for the binomial distribution is

$$\begin{aligned} B(k;p,n):F(k)\ &=\ I_{1-p}(n-k,1+k)\ = \\ &=\ 1-I_p(1+k,n-k)\ =\ (n-k)\binom{n}{k}\int_0^{1-p}t^{n-k-1}(1-t)^k\ \mathrm{dt} \end{aligned} \quad (8.33)$$

Herein $I_x(a,b)$ is the *regularized incomplete beta function*:[15]

$$I_x(a,b)\ =\ \sum_{j=a}^{a+b+1}\frac{(a+b-1)!}{j!\,(a+b-1-j)!}\,x^j\,(1-x)^{a+b-1-j}\ .$$

The de Moivre-Laplace theorem provides the proof for the convergence of the binomial distribution to a normal distribution with $\hat{\mu}=np$ and variance

[15]The *beta function* or the Euler integral of the first kind is defined by

$$B(x,y)\ =\ \int_0^1 t^{(}x-1)\,(1-t)^{(}y-1)\,\mathrm{dt}\ \ \text{for}\ \ \big(\Re(x),\Re(y)\big)>0\ ,$$

and generalized to the *incomplete beta function*

$$B(z;x,y)\ =\ \int_0^z t^{(}x-1)\,(1-t)^{(}y-1)\,\mathrm{dt}\ \ \text{and}\ \ I_z(x,y)\ =\ \frac{B(z;x,y)}{B(x,y)}\ .$$

The *regularized incomplete beta function* is thus obtained through a kind of normalization.

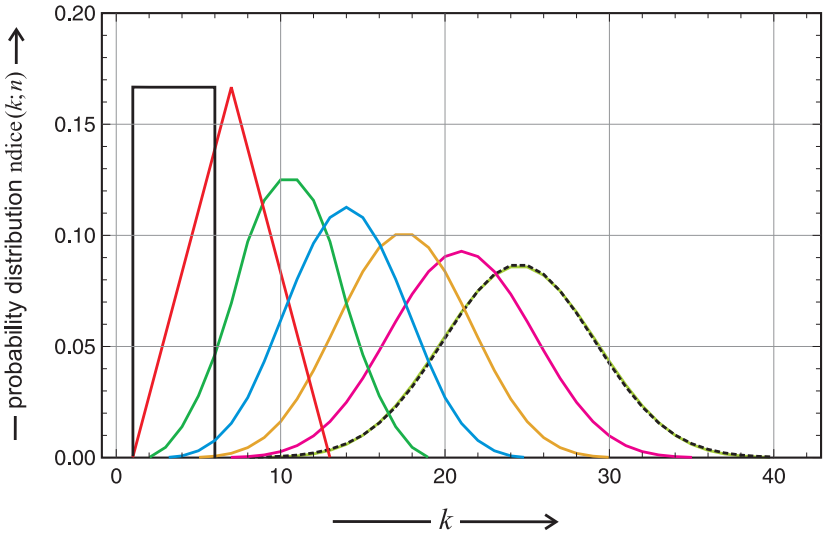


Figure 8.10: **Convergence of a probability mass function to the normal density.** The series starts with a pulse function $f(k) = 1/6$ for $i = 1, \ldots, 6$ $(n = 1)$, then comes a tent function $(n = 2)$ and then follows the gradual approach of an normal distribution, $(n = 3, 4, \ldots)$. For $n = 7$ we show the comparison with a fitted normal distribution (broken black curve). Choice of parameters: $s = 6$ and $n = 1$ (black), 2 (red), 3 (green), 4 (blue), 5 (yellow), 6 (magenta), and 7 (chartreuse).

$\sigma^2 = np(1-p) = npq$ in the limit $n \to \infty$. In particular, the theorem states

$$\lim_{n\to\infty} \frac{\binom{n}{k} p^k q^{n-k}}{\frac{\exp\left(-(k-np)^2/(2npq)\right)}{\sqrt{2\pi npq}}} \;=\; 1 \quad \text{for} \quad p+q=1, p>0, q>0\ , \tag{8.34}$$

as n becomes larger and larger k approaches a continuous variable and the binomial distribution becomes a normal distribution.

The second example deals with the extension of the rolling-dice problem to n dice. The probability of a score of k points can be calculated by means of combinations:

$$f_{s,n}(k) \;=\; \frac{1}{s^n} \sum_{i=0}^{\lfloor \frac{k-n}{s} \rfloor} (-1)^i \binom{n}{i} \binom{k - s\,i - 1}{n-1} \tag{8.35}$$

The results for small values of n and ordinary dice $(s = 6)$ are illustrated in Fig. 8.10. The convergence to a continuous probability density is nicely

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

The central limit theorem (CLT) is the generalization of the two examples shown here to arbitrary probability distributions.[16] In the traditional form it is expressed as:

Central limit theorem. Let $\{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n\}$ be a random sample of size n that is obtained as a sequence of independent and identically distributes random variables drawn from distributions with expectation values, $E(\mathcal{X}) = \mu$ and variances, $\sigma^2(\mathcal{X}) = \sigma^2$. Then the sample average of these random variables is defined by $\mathcal{S}_n := (\mathcal{X}_1 + \mathcal{X}_2 + \ldots + \mathcal{X}_n)/n$. The central limit theorem states that as n gets larger the distribution of the difference between $\mathcal{S}_n$ and its limit μ multiplied by $\sqrt{n}$, i.e. $\sqrt{n}(\mathcal{S}_n - \mu)$, approximates a normal distribution with mean zero and variance σ^2. For fixed large n this is tantamount to the statement that for fixed large n the distribution $\mathcal{S}_n$ is close to the normal distribution with expectation value μ and variance σ^2/n. In other words the distribution of $\sqrt{n}(\mathcal{S}_n - \mu)$ approaches normality regardless of the distribution of the individual $\mathcal{X}_i$'s.

The *law of large numbers* follows as a straightforward consequence of the central limit theorem. Its main message says that for a sufficiently large number of independent events the statistical errors will vanish by summation and the mean of any finite sample converges to the (exact) expectation values and higher moments:

$$\begin{aligned} \hat{m} &= \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \lim_{n\to\infty} \hat{m} = \hat{\mu}\ , \quad \text{and} \\ m_2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{m})^2 \quad \text{and} \quad \lim_{n\to\infty} m_2 = \mathrm{var}(x) = \sigma^2(x), \end{aligned} \tag{8.36}$$

where n denotes here the sample size. The sample mean $\hat{m}$ and the sample variance m_2 converge to the expectation value $\hat{\mu}$ and the variance σ^2 in the

[16] For some distributions, which have no defined moments like the Cauchy-Lorentz distribution neither the central limit theorem nor the law of large numbers are fulfilled.

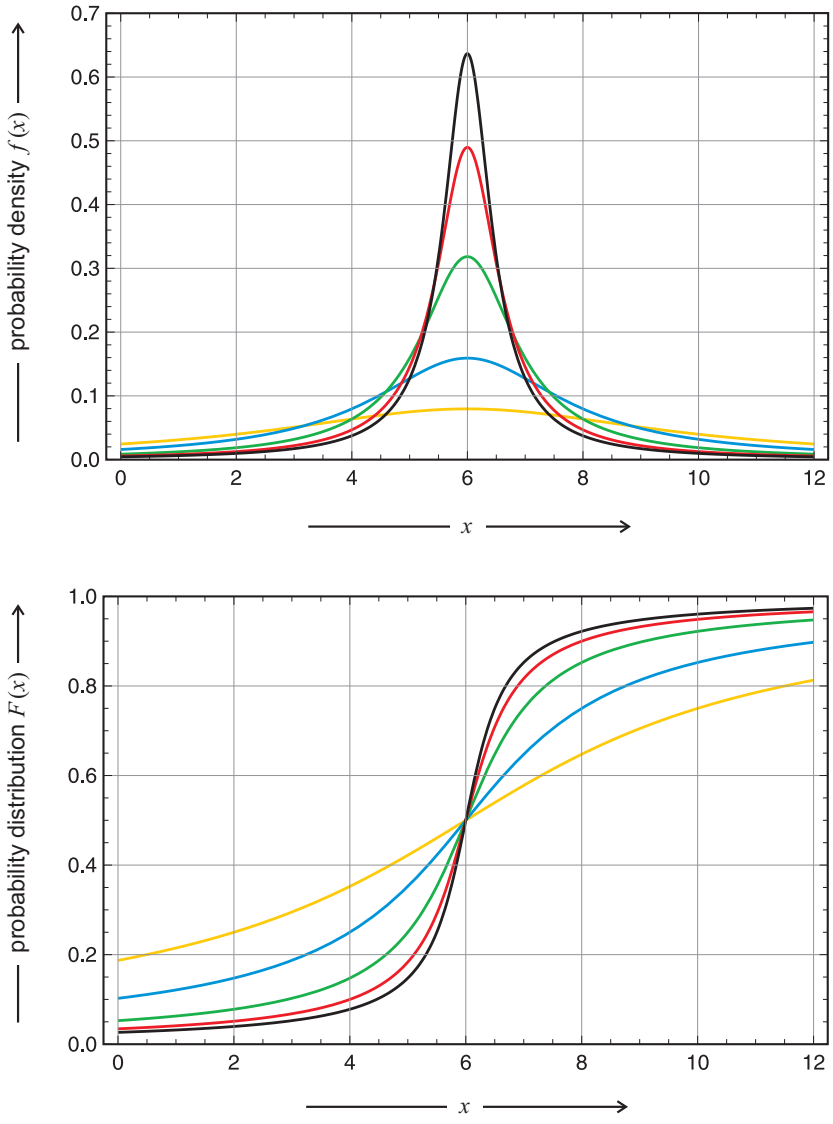


Figure 8.11: **Cauchy-Lorentz density and distribution.** In the plots the Cauchy-Lorentz distribution, $\mathcal{C}(x_0, \gamma)$, is shown in from of the probability density $f(x) = \gamma \;/\; \Big(\pi\big((x - x_0)^2 + \gamma^2\big)\Big)$ and the probability distribution $F(x) = \frac{1}{2} + \arctan\Big((x - x_0)/\gamma\Big) \;/\; \pi$. Choice of parameters: $x_0 = 6$ and $\gamma = 0.5$ (black), 0.65 (red), 1 (green), 2 (blue) and 4 (yellow).

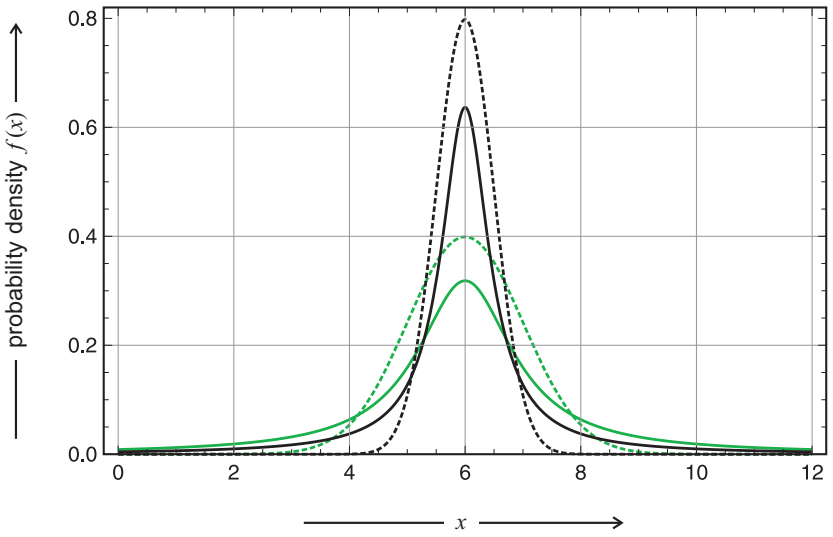


Figure 8.12: **Comparison of Cauchy-Lorentz and normal density.** The plots compare the Cauchy-Lorentz density, $\mathcal{C}(x_0, \gamma)$, and the normal density $\mathcal{N}(\mu, \sigma^2)$. In the flanking regions the normal density decays to zero much faster than the Cauchy-Lorentz density, and this is the cause of the abnormal behavior of the latter. Choice of parameters: $x_0 = \mu = 6$ and $\gamma = \sigma^2 = 0.5$ (black), 1 (green).

limit of infinite sample size. Thus the law of large numbers provides the basis for the conventional assumption of convergence with increasing sample size in mathematical statistics.

The Cauchy-Lorentz distribution. The Cauchy-Lorentz distribution, $\mathcal{C}(x_0, \gamma)$, named after the French mathematician Augustin Cauchy and the Dutch physicist Hendrik Lorentz is often addressed as the canonical example of a pathological distribution, because all its moments are either undefined like all odd moments or diverge like for example the raw second moment, $\hat{\mu}_2 = \infty$. Clearly, it has no moment generating function. Nevertheless, it has important applications, for example it is the solution to the differential equation describing forced resonance, it is used in spectroscopy to describe the spectral lines that show homogeneous broadening, and it is the probabilistic basis for a class of highly irregular stochastic processes with the Cauchy process being its most prominent example (see subsection 8.2.2).

The probability density function and cumulative distribution function are of the form

$$\mathcal{N}(x; x_0, \gamma): \quad f(x) = \frac{1}{\pi} \frac{\gamma}{(x - x_0)^2 + \gamma^2} , \tag{8.37}$$

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x - x_0}{\gamma}\right) . \tag{8.38}$$

In Fig. 8.11 the Cauchy-Lorentz density and distribution are shown for different parameter values. At a first glance the family of curves looks very similar to the analogous family of the normal distribution. A more close inspection, however, shows the flanking regions are much flatter in the Cauchy-Lorentz case (Fig. 8.12), and the slow convergence towards zero is the ultimate cause for the pathological values of the moments.

8.2 Stochastic processes

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 [28]. Although Brown himself had already demonstrated that *Brownian motion* is not caused by some (mysterious) biological effect, its origin remained kind of a mystery until Albert Einstein [74], and independently by Marian von Smoluchowski [296], published a satisfactory explanation in 1905 and 1906, respectively, which contained two main points:

(i) The motion is caused by highly frequent impacts on the pollen grain of the steadily moving molecules in the liquid in which it is suspended.

(ii) The motion of the molecules in the liquid is so complicated in detail that its effect on the pollen grain can only be described probabilistically in terms of frequent statistically independent impacts.

In particular, Einstein showed that the number of particles per unit volume, $\rho(x,t)$,[17] fulfils the already known differential equation of diffusion,

$$\frac{\partial\rho}{\partial t} = D\frac{\partial^2\rho}{\partial x^2} \quad \text{with the solution} \quad \rho(x,t) = \frac{N}{\sqrt{4\pi D}}\,\frac{\exp\Bigl(-x^2/(4Dt)\Bigr)}{\sqrt{t}}\;,$$

where N is the total number of particles. From the solution of the diffusion equation Einstein computes the square root of the mean square displacement, λ_x, the particle experiences in x-direction:

$$\lambda_x = \sqrt{\bar{x}^2} = \sqrt{2Dt}\;.$$

Einstein's treatment is based on discrete time steps and thus contains an approximation – that is not well justified – but it represents the first analysis based on a probabilistic concept of a process that is comparable to the current theories and we may consider Einstein's paper as the beginning of stochastic modeling. Brownian motion was indeed the first completely random process that became accessible to a description that was satisfactory by the standards

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

of classical physics. Thermal motion as such had been used previously as the irregular driving force causing collisions of molecules in gases by James Clerk Maxwell and Ludwig Boltzmann. The physicists in the second half of the nineteenth century, however, were concerned with molecular motion only as it is required to describe systems in the thermodynamic limit. They derived the desired results by means of global averaging statistics. For scholars interested in a mathematically rigorous history of stochastic processes we recommend the comprehensive review by Subrahmanyan Chandrasekhar [35].

Systems evolving probabilistically in time can be rigorously described and modeled in mathematical terms by *stochastic processes*. More precisely, we postulate the existence of a time dependent random variable or random vector, $\mathcal{X}(t)$ or $\vec{\mathcal{X}}(t) = \big(\mathcal{X}_1(t), \mathcal{X}_2(t), \ldots, \mathcal{X}_n(t)\big)$, respectively. Depending on the nature of the process a discrete or a continuous variable may be appropriate for modeling and we shall distinguish and consider both cases here: (i) the simpler discrete case, in which the variables refer, for example, to particle numbers

$$X_k(t) \;=\; P\Big(\mathcal{X}(t) = k\Big) \;\text{ with }\; k \in \mathbb{N}^0 \;, \tag{8.39}$$

and (ii) the continuous probability density case, where the variable are concentrations

$$dF(x,t) \;=\; f(x,t)\,dx \;=\; P\Big(x \leq \mathcal{X}(t) \leq x + dx\Big) \;\text{ with }\; x \in \mathbb{R}^1 \;. \tag{8.40}$$

The latter case, of course, requires that $dF(x,t)$ is differentiable. In both cases a *trajectory* is understood as a recording of the particular values of $\mathcal{X}$ at certain times:

$$\mathcal{T} \;=\; \Big((x_1,t_1),(x_2,t_2),(x_3,t_3),\cdots,(y_1,\tau_1),(y_2,\tau_2),\cdots\Big) \;. \tag{8.41}$$

As the repeated measurement of a stochastic variable yields different values, the repetition of a stochastic process through reproducing an experiment yields a different trajectory and the full description of the process requires knowledge on the time course of the probability distribution. In experiments and in computer simulation the probability distribution is usually derived

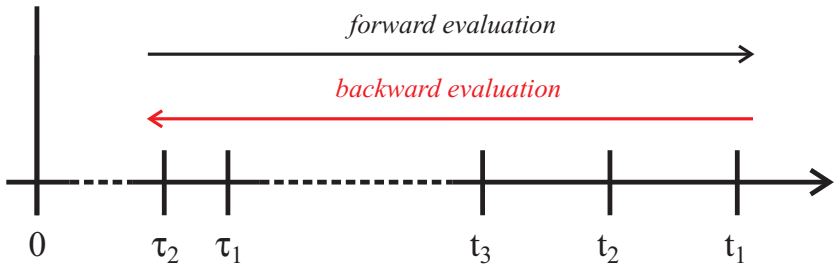


Figure 8.13: **Time order in modeling stochastic processes.** Time is progressing from left to right and the most recent event is given by the rightmost recording at time t_1. The Chapman-Kolmogorov equation describing stochastic processes comes in two forms: (i) the forward equation predicting the future form past and present and (ii) the backward equation, which extrapolates back in time from present to past.

through superposition of many individual trajectories (see subsection 9.3). Although it is not essential for the application of probability theory but definitely required for comparison with experiments, we shall assume here that the recorded values are time ordered with the oldest values on the rightmost position and the most recent values at the latest entry on the left-hand side (Fig. 8.13):[18]

$$t_1 \geq t_2 \geq t_3 \geq \cdots \geq \tau_1 \geq \tau_2 \geq \cdots \ .$$

A trajectory thus is a time ordered sequence of doubles (x, t).

In order to model evolution or other dynamical phenomena based on chemical or biochemical kinetics by stochastic processes we postulate a population of molecular species $\mathcal{P} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n\}$, which are quantitatively described by a time dependent random vector $\vec{\mathcal{X}}(t)$ as defined above. In modeling evolution the species $\mathbf{X}_j$ are various individual agents that are capable of reproduction and mutation. In the deterministic approach (chapter 4) we postulated a space of all possible genotypes, the sequence space $\mathcal{Q}_l^{(\kappa)}$, and we chose it as a theoretical reference for evolution. In case of stochastic

[18] Time is measured with respect to some reference point on the right-hand side, and we remark that the *stochastic time axis* runs in the opposite direction of the conventional time axis in plots.

processes it is much more appropriate to restrict the population to the actually present variants and to introduce a new variable whenever a mutation leads to a genotype that was not present in the population before. Then, the size of the population support $\Sigma(\mathcal{P})$ is changing and can be considered as a function of time, $|\Sigma(\mathcal{P})| = n_\Sigma(t)$. Variable population size N and variable species diversity n_Σ have great impact on the evolutionary process. For example bottle necks and broad regions shape evolution as will be discussed later in section 10.2 [101, 246, 247].

8.2.1 Markov and other simple stochastic processes

A stochastic process, as we shall assume, is determined by a set of joint probability densities the existence and analytical form of which is presupposed.[19] The probability density encapsulates the physical nature of the process and contains all parameters and data on external conditions and hence we can assume that they determine the system completely:

$$p\big(x_1, t_1; x_2, t_2; x_3, t_3; \cdots\big) \ . \tag{8.42}$$

By the phrase "*determine completely*" we mean that no additional information is required for a description of the progress in terms of the time ordered series (8.41) 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 (8.42) is straightforward by application of *marginal densities*. For the discrete case the result is obvious

$$P(\mathcal{X} = x_1) \ = \ p(x_1, *) \ = \sum_{x_k \neq x_1} p\left(x_1, t_1; x_2, t_2; x_3, t_3; \cdots; x_n, t_n; \cdots\right) \ .$$

[19] The joint density p (8.14) is defined by $\text{Prob}\left(\mathcal{X} = x_i, \mathcal{Y} = y_j\right) = p(x_i, y_j)$ or $\text{Prob}\left(\mathcal{X} \leq x, \mathcal{Y} \leq y\right) = \int_{-\infty}^{x}\int_{-\infty}^{x} f(u, v)\,\mathrm{d}u\,\mathrm{d}v$ in the continuous case. Since we are always dealing with doubles (x, t) we modify the notation slightly and separate individual doubles by semicolons: $\cdots; x_k, t_k; x_{k+1}, t_{k+1}; \cdots$.

and in the continuous case we obtain

$$P(\mathcal{X}_1 = x_1 \in [a,b]) \;=\; \int_a^b \mathrm{d}x_1 \iiint_{-\infty}^{\infty} \mathrm{d}x_2\mathrm{d}x_3\cdots\mathrm{d}x_n\cdots \\ p\,(x_1,t_1;x_2,t_2;x_3,t_3;\cdots;x_n,t_n;\cdots)$$

Time ordering allows for the formulation of predictions on future values from the known past in terms of conditional probabilities:

$$p\,(x_1,t_1;x_2,t_2;\cdots|\,y_1,\tau_1;y_2,\tau_2,\cdots) \;=\; \frac{p\,(x_1,t_1;x_2,t_2;\cdots;y_1,\tau_1;y_2,\tau_2,\cdots)}{p\,(y_1,\tau_1;y_2,\tau_2,\cdots)}\,,$$

with $t_1 \geq t_2 \geq \cdots \geq \tau_1 \geq \tau_2 \geq \cdots$. In other words, we may compute $\{(x_1,t_1),(x_2,t_2),\cdots\}$ from known $\{(y_1,\tau_1),(y_2,\tau_2),\cdots\}$.

Three simple stochastic processes will be discussed here: (i) the factorizable process with probability densities that are independent of the event 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.

The simplest class of stochastic processes is characterized by complete independence of events,

$$p\,(x_1,t_1;x_2,t_2;x_3,t_3;\cdots) \;=\; \prod_i p\,(x_i,t_i)\;, \tag{8.43}$$

which 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 where the probability densities are also independent of time: $p\,(x_i,t_i) = p\,(x_i)$, and then we have

$$p\,(x_1,t_1;x_2,t_2;x_3,t_3;\cdots) \;=\; \prod_i p\,(x_i)\;. \tag{8.44}$$

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 – and then the product is replaced by $p(x)^n$.

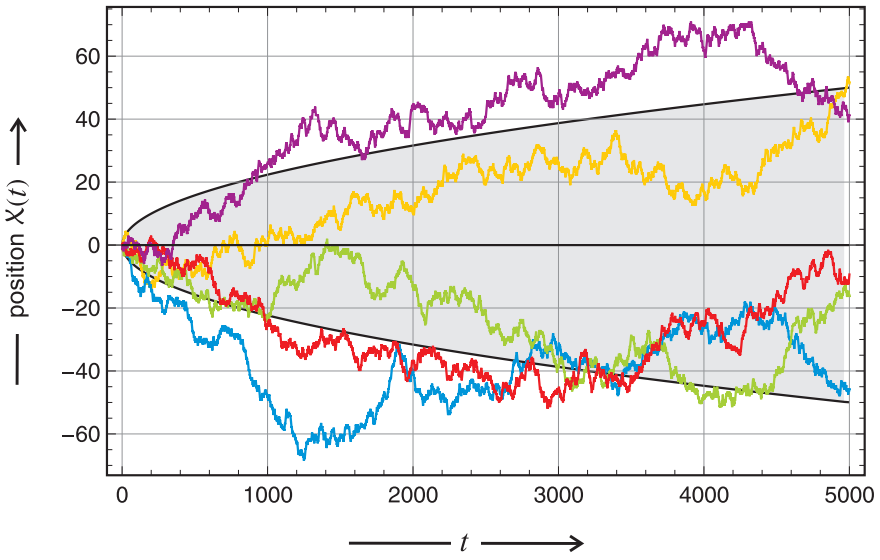


Figure 8.14: **The one-dimensional random walk.** The one-dimensional random is shown as an example of a martingale. Five trajectories were calculated with different seeds for the random number generator. The expectation value $E\big(\mathcal{X}(t)\big) = x_0 = 0$ is constant and the variance grows linearly with time $\sigma^2\big(\mathcal{X}(t)\big) = 2dt$, Accordingly, the standard deviation grows with $\sqrt{t}$. The three black lines in the figure correspond to E and $E \pm \sigma(t)$, and the grey area represent the confidence interval of 68,2 %. Choice of parameters: $\vartheta = \frac{1}{2}$; random number generator: *Mersenne Twister*; seeds: 491 (yellow), 919 (blue), 023 (green), 877 (red), 127 (violet).

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. The conditional mean value of the random variable $\mathcal{X}(t)$ provided $\mathcal{X}(t_0) = x_0$ is defined as

$$E\big(\mathcal{X}(t)|(x_0,t_0)\big) \doteq \int \mathrm{d}x\, p(x,t|x_0,t_0) \ .$$

In a martingale the conditional mean is simple given by

$$E\big(\mathcal{X}(t)|(x_0,t_0)\big) = x_0 \ . \tag{8.45}$$

The mean value at time t is identical to the initial value of the process (Fig. 8.14). The martingale property is rather strong and restrictive and

applies only to relatively few cases to which we shall refer only in some specific situations.

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 a *càdlàg* adapted process with bounded variation[20]

$$\mathcal{X}(t) \;=\; \mathcal{M}(t) + \mathcal{A}(t)$$

A local martingale is a stochastic process that satisfies locally the martingale property (8.45) 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 or nonanticipating process is a process that *cannot see into the future*. An informal interpretation [310, section II.25] would say: A stochastic process $\mathcal{X}(t)$ is adapted iff for every realization and for every time t, $\mathcal{X}(t)$ is known at time t and not before.

Another simple concept assumes that knowledge of the present only is sufficient to predict the future. It is realized in *Markov processes* named after the Russian mathematician Andrey Markov[21] and can formulated easily in terms of conditional probabilities:

$$p\,(x_1,t_1;x_2,t_2;\cdots\,|\,y_1,\tau_1;y_2,\tau_2,\cdots) \;=\; p\,(x_1,t_1;x_2,t_2;\cdots\,|\,y_1,\tau_1)\;. \quad (8.46)$$

In essence, the Markov condition expresses independence of the history of the process prior to time τ_1 or in other words and said more sloppily: "A Markov process has no memory and the future is completely determined by

[20]The property *càdlàg* is an acronym from French for "*continue à droite, limites à gauche*". It is a common property in statistics (section 8.1).

[21]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 [74] and Marian von Smoluchowski [296] in their derivation of the diffusion process.

the presence." In particular, we have

$$p\,(x_1,t_1;x_2,t_2;y_1,\tau_1) \;=\; p\,(x_1,t_1|\,x_2,t_2)\,p\,(x_2,t_2|\,y_1,\tau_1)\;.$$

Any arbitrary joint probability can be simply expressed as products of conditional probabilities:

$$\begin{aligned} &p\,(x_1,t_1;x_2,t_2;x_3,t_3;\cdots;x_n,t_n) \;= \\ &=\; p\,(x_1,t_1|\,x_2,t_2)\,p\,(x_2,t_2|\,x_3,t_3)\cdots p\,(x_{n-1},t_{n-1}|\,x_n,t_n)\,p\,(x_n,t_n) \end{aligned} \tag{8.46'}$$

under the assumption of time ordering $t_1 \geq t_2 \geq t_3 \geq \ldots t_{n-1} \geq t_n$.

8.2.2 Continuity and the Chapman-Kolmogorov equation

Prior to the discussion of the mathematical approach to model step-processes, continuous processes, and combinations of both we shall consider a classification of stochastic processes. Assuming a discrete representation of particle numbers as in Equ. (8.39) the probability functions of the stochastic variable $\mathcal{X}(t)$ can change only in steps or jumps. A continuous stochastic variable $\mathcal{X}(t)$ with a suitable probability function, however, need not exclude the occurrence of jumps. Accordingly, we expect to be dealing with several classes of processes: (i) pure deterministic or *drift processes*, (ii) pure driftless and jump free *diffusion processes*, (iii) pure jump processes, (iv) a combination of (i) and (ii) being processes with drift and diffusion, and (v) a combination of (i), (ii), and (iii) and others. At first we introduce the notion of continuity in stochastic processes be means of two examples, the Wiener and the Cauchy process. Then, we shall discuss a very general equation for the description of stochastic processes, which contains all above mentioned processes as special cases.

Continuity in stochastic processes. The condition of continuity in Markov processes requires a more detailed discussion. The process goes from position z at time t to position x at time $t+\Delta t$. Continuity of the process implies that the probability of x to be finitely different from z goes to zero faster than Δt in the limit $\lim \Delta t \to 0$:

$$\lim_{\Delta t\to 0}\,\frac{1}{\Delta t}\int_{|x-z|>\varepsilon} dx\,p\,(x,t+\Delta t|z,t) \;=\; 0\;, \tag{8.47}$$

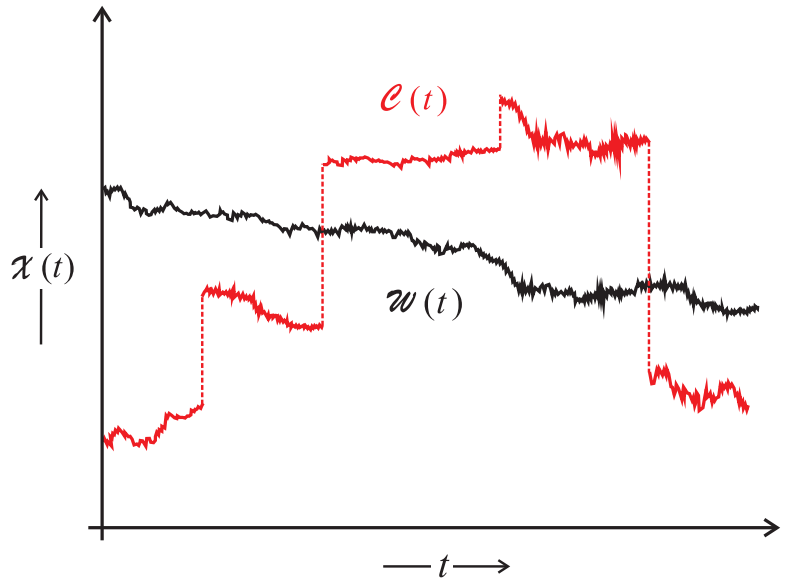


Figure 8.15: **Continuity in Markov processes.** Continuity is illustrated by means of two stochastic processes of the random variable $\mathcal{X}(t)$, the Wiener process $\mathcal{W}(t)$ (8.48) and the Cauchy process $\mathcal{C}(t)$ (8.49). The Wiener process describes Brownian motion and is continuous but almost nowhere differentiable. The even more irregular Cauchy process is wildly discontinuous.

and this uniformly in z, t, and Δt. In other words, the difference in probability as a function of $|\,x-z|$ converges sufficiently fast to zero and hence no jumps occur in the random variable $\mathcal{X}(t)$.

Two illustrative examples for the analysis of continuity are chosen and sketches in Fig. 8.15: (i) the Einstein-Smoluchowski solution of Brownian motion that is a continuous version of the one-dimensional random walk shown in Fig. 8.14,[22] which leads to a normally distributed probability,

$$p(x,t+\Delta t|z,t) \;=\; \frac{1}{\sqrt{4\pi D\Delta t}}\;\exp\left(-\frac{(x-z)^2}{4D\Delta t}\right)\;, \tag{8.48}$$

and (ii) the so-called Cauchy process following the Cauchy-Lorentz distribu-

[22]Later on we shall discuss the continuous version of this stochastic process in more detail and call it a *Wiener process*.

tion,

$$p(x, t+\Delta t|z, t) \;=\; \frac{\Delta t}{\pi}\,\frac{1}{(x-z)^2+\Delta t^2}\;. \tag{8.49}$$

In case of the Wiener process we exchange the limit and the integral, introduce $\vartheta = (\Delta t)^{-1}$, perform the limit $\vartheta \to \infty$, and have

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|x-z|>\varepsilon} \mathrm{d}x\,\frac{1}{\sqrt{4\pi D}}\,\frac{1}{\sqrt{\Delta t}}\,\exp\left(-\frac{(x-z)^2}{4D\Delta t}\right) \;=$$

$$=\int_{|x-z|>\varepsilon} \mathrm{d}x\,\lim_{\Delta t\to 0}\frac{1}{\Delta t}\,\frac{1}{\sqrt{4\pi D}}\,\frac{1}{\sqrt{\Delta t}}\,\exp\left(-\frac{(x-z)^2}{4D\Delta t}\right) \;=$$

$$=\int_{|x-z|>\varepsilon} \mathrm{d}x\,\lim_{\vartheta\to\infty}\frac{1}{\sqrt{4\pi D}}\,\frac{\vartheta^{3/2}}{\exp\left(\frac{(x-z)^2}{4D}\,\vartheta\right)}\,, \quad \text{where}$$

$$\lim_{\vartheta\to\infty}\frac{\vartheta^{3/2}}{1\;+\;\frac{(x-z)^2}{4D}\cdot\vartheta\;+\;\frac{1}{2!}\left(\frac{(x-z)^2}{4D}\right)^2\cdot\vartheta^2\;+\;\frac{1}{3!}\left(\frac{(x-z)^2}{4D}\right)^3\cdot\vartheta^3\;+\;\ldots} \;=\; 0\;.$$

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 curve of Brownian motion [28] is indeed extremely irregular since it is nowhere differentiable (Fig. 8.15).

In the second example, the Cauchy process, we exchange limit and integral as in case of the Wiener process, and perform the limit $\Delta t \to 0$:

$$\lim_{\Delta t\to 0}\frac{1}{\Delta t}\int_{|x-z|>\varepsilon} \mathrm{d}x\,\frac{\Delta t}{\pi}\,\frac{1}{(x-z)^2+\Delta t^2} \;=$$

$$=\int_{|x-z|>\varepsilon} \mathrm{d}x\,\lim_{\Delta t\to 0}\frac{1}{\Delta t}\,\frac{\Delta t}{\pi}\,\frac{1}{(x-z)^2+\Delta t^2} \;=$$

$$=\int_{|x-z|>\varepsilon} \mathrm{d}x\,\lim_{\Delta t\to 0}\frac{1}{\pi}\,\frac{1}{(x-z)^2+\Delta t^2} \;=\; \int_{|x-z|>\varepsilon}\frac{1}{\pi(x-z)^2}\,\mathrm{d}x \;\neq\; 0\;.$$

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 the curve for the Cauchy-process is also irregular but even discontinuous.

Both processes, as required for consistency, fulfill the relation

$$\lim_{\Delta t\to 0} p\,(x, t+\Delta t|\,z, t) \;=\; \delta(x-z)\;,$$

where $\delta(\cdot)$ is the so-called delta-function.[23]

Chapman-Kolmogorov equation. From joint probabilities follows that summation of all mutually exclusive events of one kind eliminates the corresponding variable:

$$\sum_i P(A \cap B_i \cap C \cdots) \;=\; P(A \cap C \cdots) \;,$$

where the subsets B_i fulfil the conditions

$$B_i \cap B_j \;=\; \emptyset \;\text{ and }\; \bigcup_i B_i \;=\; \Omega \;.$$

Applied to stochastic processes we find by the same token

$$p\,(x_1,t_1) \;=\; \int dx_2\, p\,(x_1,t_1;x_2,t_2) \;=\; \int dx_2\, p\,(x_1,t_1|\,x_2,t_2)\, p\,(x_2,t_2) \;.$$

Extension to three events leads

$$\begin{aligned} p\,(x_1,t_1|x_3,t_3) \;&=\; \int dx_2\, p\,(x_1,t_1;x_2,t_2|\,x_3,t_3) \;= \\ &=\; \int dx_2\, p\,(x_1,t_1|\,x_2,t_2;x_3,t_3)\, p\,(x_2,t_2|\,x_3,t_3) \;. \end{aligned}$$

For $t_1 \geq t_2 \geq t_3$ and making use of the Markov assumption we obtain the *Chapman-Kolmogorov equation*, which is named after the British geophysicist and mathematician Sydney Chapman and the Russian mathematician Andrey Kolmogorov:

$$p\,(x_1,t_1|x_3,t_3) \;=\; \int dx_2\, p\,(x_1,t_1|\,x_2,t_2)\, p\,(x_2,t_2|\,x_3,t_3) \;. \tag{8.50}$$

In case of discrete random variable $\mathcal{N} \in \mathbb{N}^0 = 0,1,\ldots,k,\ldots$ defined on the integers we replace the integral by a sum and obtain

$$P(k_1,t_1|\,k_3,t_3) \;=\; \sum_{k_2} P(k_1,t_1|\,k_2,t_2)\; P(k_2,t_2|\,k_3,t_3) \;. \tag{8.51}$$

[23] 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, [244, pp.585-590] and [241, pp.38-42].

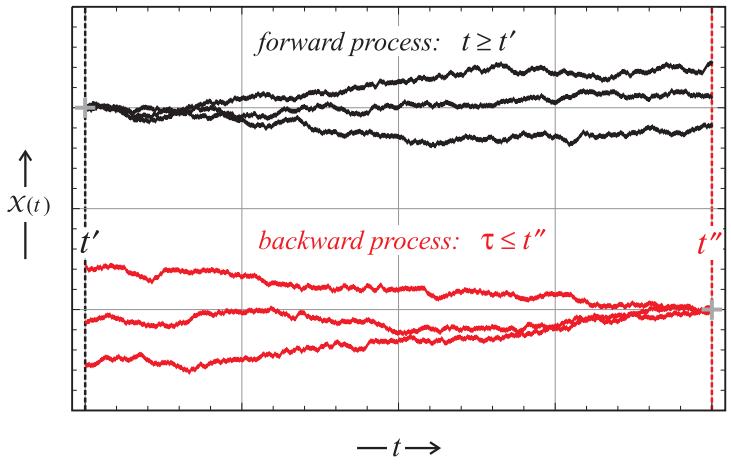


Figure 8.16: **Illustration of forward and backward equations.** The forward differential Chapman-Kolmogorov equation starts from an initial condition corresponding to the sharp distribution $\delta(\boldsymbol{y} - \boldsymbol{z})$, $(\boldsymbol{y}, t')$ is fixed (black), and the probability density unfolds with time $t \geq t'$ (black). It is well suited for the description of actual experimental situations. The backward equation, although somewhat more convenient and easier to handle from the mathematical point of view, is less suited to describe typical experiments and commonly applied to first passage time or exit problems. Here, $(\boldsymbol{x}, t'')$ is held constant (red) and the time dependence of the probability density corresponds to samples unfolding into the past, $\tau \leq t''$ (red). The *initial condition*, $\delta(\boldsymbol{y} - \boldsymbol{z})$, is in this case replaced by a *final condition*, $\delta(\boldsymbol{z} - \boldsymbol{x})$, represented by a sharp distribution.

The Chapman-Kolmogorov equation can be interpreted in two different ways and the corresponding implementations are known as *forward* and *backward* equation (Fig. 8.16). In the forward equation the double (x_3, t_3) is considered to be fixed and (x_1, t_1) represents the variable $x_1(t_1)$, and the time t_1 proceeding in positive direction. The backward equation is exploring the past of a a given situation: the double (x_1, t_1) is fixed and (x_3, t_3) is propagating backwards in time. The forward equation is better suited to describe actual processes, whereas the backward equation is the appropriate tool to compute the evolution towards given events, for example first passage times. In order to discuss the structure of solutions of Eqs. (8.50) and (8.51), we shall present the equations in differential form (For a derivation see [107]).

The differential version of the Chapman-Kolmogorov equation is based on the continuity condition already discussed for the Wiener and the Cauchy process. In the derivation of the differential form the general equations (8.50) and (8.51) have to be partitioned with respect to differentiability conditions corresponding either to continuous motion under generic conditions or to discontinuous motion. This partitioning is based on the following conditions for all $\varepsilon > 0$: (8.52)

(i) $\lim_{\Delta t\to 0} \frac{1}{\Delta t}\, p(\boldsymbol{x}, t+\Delta t|\boldsymbol{z}, t)\, t \;=\; W(\boldsymbol{x}|\boldsymbol{z}, t)$, uniformly in $\boldsymbol{x}$, $\boldsymbol{z}$, and t for $|\boldsymbol{x}-\boldsymbol{z}| \geq \varepsilon$,

(ii) $\lim_{\Delta t\to 0} \frac{1}{\Delta t} \int_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} \mathrm{d}\boldsymbol{x}\,(x_i - z_i)\, p(\boldsymbol{x}, t+\Delta t|\boldsymbol{z}, t) \;=\; A_i(\boldsymbol{z}, t) \,+\, \mathcal{O}(\varepsilon)$, uniformly in $\boldsymbol{z}$, and t, and

(iii) $\lim_{\Delta t\to 0} \frac{1}{\Delta t} \int_{|\boldsymbol{x}-\boldsymbol{z}|<\varepsilon} \mathrm{d}\boldsymbol{x}\,(x_i - z_i)(x_j - z_j)\, p(\boldsymbol{x}, t+\Delta t|\boldsymbol{z}, t) \;=$
$= B_{ij}(\boldsymbol{z}, t) \,+\, \mathcal{O}(\varepsilon)$, uniformly in $\boldsymbol{z}$, and t.

where x_i, x_j, z_i, and z_j, refer to particular components of the vectors $\boldsymbol{x}$ and $\boldsymbol{z}$, respectively. It is important to notice that all higher-order coefficients of motion C_{ijk}, defined in analogy to A_i in (ii) or B_{ij} in (iii), must vanish by symmetry considerations [107, p.47-48]. In essence, the three terms in Equ. (8.52) refer to different behavior with respect to continuity: (i) (deterministic) drift, (ii) diffusion in the sense of a Wiener process, and (iii) a jump process, with the attributes (i) continuous and differentiable, (ii) continuous and nowhere differentiable, and (iii) discontinuous, respectively.

The partial differential equation for the evolution of the probability distribution has been called *differential Chapman-Kolmogorov equation* by Crispin Gardiner. Neglect of surface terms at the boundary of the domain of $\boldsymbol{z}$ leads to the expression:

$$\frac{\partial p\,(\boldsymbol{z}, t\,|\,\boldsymbol{y}, t')}{\partial t} = -\sum_i \frac{\partial}{\partial z_i}\Big(A_i(\boldsymbol{z}, t)\, p\,(\boldsymbol{z}, t|\,\boldsymbol{y}, t')\Big) + \tag{8.53a}$$

$$+\,\frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial z_i \partial z_j}\Big(B_{ij}(\boldsymbol{z}, t)\, p\,(\boldsymbol{z}, t|\,\boldsymbol{y}, t')\Big) + \tag{8.53b}$$

$$+\int \mathrm{d}\boldsymbol{x}\,\Big(W(\boldsymbol{z}\,|\boldsymbol{x}, t)\, p\,(\boldsymbol{x}, t|\boldsymbol{y}, t') - W(\boldsymbol{x}\,|\boldsymbol{z}, t)\, p\,(\boldsymbol{z}, t|\boldsymbol{y}, t')\Big). \tag{8.53c}$$

Under certain conditions, $\boldsymbol{A}(\boldsymbol{x},t)$ real, $\mathrm{B}(\boldsymbol{x},t)$ positive semi-definite, and $W(\boldsymbol{x}|\,\boldsymbol{y},t)$ non-negative, a non-negative solution to to the differential Chapman-Kolmogorov equation (8.53), which fulfils the continuous Chapman-Kolmogorov equation (8.50), exists for the initial condition $p\,(\boldsymbol{z},t|\,\boldsymbol{y},t) = \delta(\boldsymbol{y}-\boldsymbol{z})$.

The nature of the different stochastic processes associated with the three terms in Equ. (8.53) is visualized by setting some parameters, $\boldsymbol{A}(\boldsymbol{z},t)$, $\mathrm{B}(\boldsymbol{z},t)$, and $W(\boldsymbol{x}\,|\boldsymbol{z},t)$ equal to zero and analyzing the remaining equation. We shall discuss here the four cases that are modeled by different equations

(i) $B=0$, $W=0$, deterministic drift process: Liouville equation,

(ii) $A=0$, $W=0$, diffusion process: drift free Fokker-Planck equation,

(iii) $W=0$, drift and diffusion process: Fokker-Planck equation, and

(iv) $A=0$, $B=0$, pure jump process: master equation.

The first three processes can be modeled very efficiently by stochastic differential equations (SDE) often called Langevin equations after the French physicist Paul Langevin [186]. Examples of SDEs will mentioned for the Wiener process and the Ornstein-Uhlenbeck process. We shall mention them only briefly here, refrain from a detailed discussion, and refer to the literature [107, 251].

Chapman-Kolmogorov backward equation. Equ. (8.53) has been derived for the forward direction and according to Fig. 8.16 a differential Chapman-Kolmogorov equation can also be derived for the backward process. Then the limit has to be taken with respect to the initial variables $(\boldsymbol{y},\tau)$ of the probability $p\,(\boldsymbol{x},t|\,\boldsymbol{y},\tau)$ and the limit of the vanishing time interval is of the form

$$\begin{aligned}
&\lim_{\Delta\tau\to 0}\frac{1}{\Delta\tau}\Big(p\,(\boldsymbol{x},t|\,\boldsymbol{y},\tau+\Delta\tau)\;-\;p(\boldsymbol{x},t|\,\boldsymbol{y},\tau)\Big)\;=\\
=\;&\lim_{\Delta\tau\to 0}\frac{1}{\Delta\tau}\int \mathrm{d}\boldsymbol{z}\,p\,(\boldsymbol{z},\tau+\Delta\tau|\,\boldsymbol{y},\tau)\Big(p\,(\boldsymbol{x},t|\,\boldsymbol{y},\tau+\Delta\tau)\;-\;p\,(\boldsymbol{x},t|\,\boldsymbol{z},\tau+\Delta\tau)\Big)\;,
\end{aligned}$$

where the Chapman-Kolmogorov equation (8.50) was used to derive the second line. The derivation of the differential from of the backward equation follows essentially the same steps as in case of the forward equation [107] and

eventually we obtain a probabilistic differential equation that is equivalent to the forward equation:

$$\frac{\partial p\,(\boldsymbol{x},t\,|\,\boldsymbol{y},\tau)}{\partial t} = -\sum_i A_i(\boldsymbol{y},\tau)\,\frac{\partial p\,(\boldsymbol{x},t|\,\boldsymbol{y},\tau)}{\partial y_i} + \tag{8.54a}$$

$$+\,\frac{1}{2}\sum_{i,j} B_{ij}(\boldsymbol{z},t)\,\frac{\partial^2 p\,(\boldsymbol{x},t|\,\boldsymbol{y},\tau)}{\partial y_i \partial y_j} + \tag{8.54b}$$

$$+\int \mathrm{d}\boldsymbol{x}\, W(\boldsymbol{z}\,|\boldsymbol{y},\tau)\,\Big(p\,(\boldsymbol{x},t|\boldsymbol{y},\tau)\;-\;p\,(\boldsymbol{z},t|\boldsymbol{z},\tau)\Big). \tag{8.54c}$$

As we have stated for the forward equation the appropriate initial condition is $p\,(\boldsymbol{x},t|\,\boldsymbol{y},t) = \delta(\boldsymbol{x}-\boldsymbol{y})$ for all t.

Liouville equation. In the limiting case of a stochastic process with vanishing probabilistic contributions, diffusion and jump processes, one obtains a differential equation that is a *Liouville equation* well known from classical mechanics:

$$\frac{\partial p(\boldsymbol{z},t|\boldsymbol{y},t')}{\partial t} \;=\; -\sum_i \frac{\partial}{\partial z_i}\Big(A_i(\boldsymbol{z},t)\,p(\boldsymbol{z},t|\boldsymbol{y},t')\Big)\;, \tag{8.55}$$

Equ. (8.55) describes a completely deterministic motion, which is a solution of the ordinary differential equation

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{dt}} \;=\; \boldsymbol{a}\big(\boldsymbol{x}(t),t\big) \;\text{ with }\; \boldsymbol{x}(\boldsymbol{y},t') \;=\; \boldsymbol{y}\;. \tag{8.56}$$

The (probabilistic) solution of the differential Equ. (8.55) with the initial condition $p(\boldsymbol{z},t'|\boldsymbol{y},t') = \delta(\boldsymbol{z}-\boldsymbol{y})$ is $p(\boldsymbol{z},t|\boldsymbol{y},t') = \delta\big(\boldsymbol{z}-\boldsymbol{x}(\boldsymbol{y},t)\big)$, or in other words $\boldsymbol{z}(0) = \boldsymbol{z}(t') = \boldsymbol{y}(t')$ represents the initial conditions of the ODE (8.56). The proof of this assertion is obtained by direct substitution [107, p.54].

$$\begin{aligned}\sum_i \frac{\partial}{\partial z_i}\Big(A_(\boldsymbol{z},t)\,\delta\big(\boldsymbol{z}-\boldsymbol{x}(\boldsymbol{y},z)\big)\Big) &= \sum_i \frac{\partial}{\partial z_i}\Big(A_i(\boldsymbol{x}(\boldsymbol{y},t),t)\,\delta\big(\boldsymbol{z}-\boldsymbol{x}(\boldsymbol{y},t)\big)\Big),\\ &= \sum_i \Big(A_i(\boldsymbol{x}(\boldsymbol{y},t),t)\;\frac{\partial}{\partial z_i}\,\delta\big(\boldsymbol{z}-\boldsymbol{x}(\boldsymbol{y},t)\big)\Big),\\ \text{and}\;\; \frac{\partial}{\partial t}\,\delta\big(\boldsymbol{z}-\boldsymbol{x}(\boldsymbol{y},t)\big) &= -\sum_i \frac{\partial}{\partial z_i}\,\delta\big(\boldsymbol{z}-\boldsymbol{x}(\boldsymbol{y},t)\big)\,\frac{\mathrm{d}x_i(\boldsymbol{y},t)}{\mathrm{dt}}\,,\end{aligned}$$

and by means of Equ. (8.56) the last two lines become equal. □

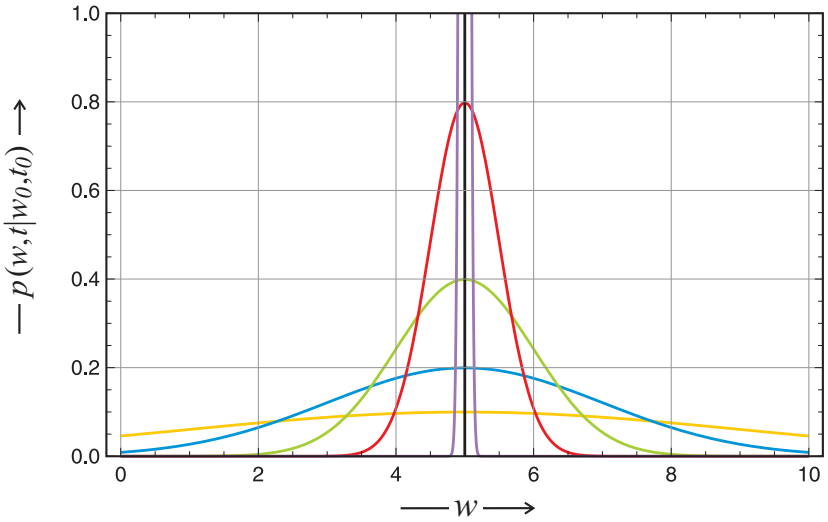


Figure 8.17: **Evolution of the probability density in the Wiener process.** In the figure we show the conditional probability density of the Wiener process which is identical with the normal distribution (Fig. 8.9),

$$p(w,t|w_0,t_0) = \exp\Big(-(w-w_0)^2/\big(2(t-t_0)\big)\Big)/\sqrt{2\pi(t-t_0)}.$$

The values used are $w_0 = 5$ and $t - t_0 = 0$ (black), 0.05 (purple), 0.5 (red), 1.0 (green), 2.0 (blue), and 4.0 (yellow). The initially sharp distribution, $p(w,t|w_0,t_0) = \delta(w-w_0)$ spreads with increasing time until it becomes completely flat in the limit $t \to \infty$.

If a particle is in a well-defined position $\boldsymbol{y}$ at time t' it will remain on the trajectory obtained by solving the corresponding ordinary differential equation (ODE). Deterministic motion can be visualized therefore as an elementary form of Markov process, which can be formulated by a drift-diffusion process with a zero diffusion matrix B. A generalization of the initial conditions is straightforward: If we are dealing with a probability distribution rather than a δ-function this distribution is traveling along the trajectory.

Wiener process. If only the diffusion matrix B has nonzero entries and all terms of $\boldsymbol{A}(\boldsymbol{z},t)$ and $W(\boldsymbol{x}|\boldsymbol{z},t)$ vanish, we are dealing with a *Wiener process* named after the American mathematician and logician Norbert Wiener. It is synonymous to (continuous) Brownian motion or *white noise* and describes among other things random fluctuations caused by thermal motion. The

trajectories in Fig. 8.14 are almost indistinguishable from those of a Wiener process because a small increment of the random walk was chosen. From the point of view of stochastic processes the probability density of a Wiener process is a solution of the differential equation[24]

$$\frac{\partial p\,(w,t|\,w_0,t_0)}{\partial t} \;=\; \frac{1}{2}\,\frac{\partial^2}{\partial w^2}\,p\,(w,t|\,w_0,t_0) \tag{8.57}$$

in one random variable, $\mathcal{W}(t)$. The probability distribution of $\mathcal{W}(t)$ is given by

$$P(\mathcal{W}(t)\leq w') = \int_{-\infty}^{w'} p\,(w,t)\,\mathrm{d}w\ .$$

Equ. (8.57) can be understood as a *Fokker-Planck equation* with drift coefficient zero and diffusion coefficient $D=1$. Fokker-Planck equations are partial differential equations named after two physicists, the Dutchman Adriaan Daniël Fokker and the German Max Planck and in one dimension, in the forward direction, and for the initial conditions $p\,(x_0,t_0)$ they are of the general form

$$\frac{\partial p\,(x,t)}{\partial t} \;=\; -\frac{\partial}{\partial x}\Big(A(x,t)\,p\,(x,t)\Big) \;+\; \frac{1}{2}\,\frac{\partial^2}{\partial x^2}\Big(B(x,t)\,p\,(x,t)\Big)\ . \tag{8.58}$$

An example of a full Fokker-Planck equation modeling the Ornstein-Uhlenbeck process will be discussed in the next paragraph. Fokker-Planck equations can be easily converted into stochastic differential equations (SDEs), which are convenient for analysis and computer simulation [107]. The probability distribution of the Fokker-Planck equation (8.58) is equivalent to the ensemble of trajectories described by

$$dx \;=\; a(x,t)\ \mathrm{dt} \;+\; b(x,t)\,dW(t)\,, \tag{8.59}$$

where $dW(t)$ corresponds to a Wiener process.[25] Integration of stochastic differential equation requires special definitions that will not be discussed here.

[24] In application in physics and chemistry diffusion occurs in space and it would be natural to use $\mathcal{X}(t)$ as stochastic variable and x as coordinate. Because of the key role of the Wiener process in stochasticity we shall use $\mathcal{W}(t)$ and w instead.

[25] The original formulation by Paul Langevin was different:

$$\tfrac{\mathrm{d}x}{\mathrm{dt}} \;=\; a(x,t) \;+\; b(x,t)\,\xi(t)\ ,$$

The solution of (8.57) for the initial condition on the conditional probability $p\,(w,t_0|\,w_0,t_0) = \delta(w-w_0)$ can be obtained by using the characteristic function

$$\phi(s,t) \;=\; \int \mathrm{d}w\, p\,(w,t|\,w_0,t_0)\, \exp(\dot{\imath} s\, w)\;,$$

which fulfils $\partial\phi(s,t)/\partial t = -\frac{1}{2}\, s^2\, \phi(s,t)$ as can be shown by applying integration in parts twice and making use of the fact that $p\,(w,t|\,w_0,t_0)$ like every probability density has to vanish in the limits $w \to \pm\infty$ and the same is true for the first partial derivative, $\partial p/\partial w$. Next we compute the characteristic function by integration:[26]

$$\phi(s,t) \;=\; \phi(s,t_0)\cdot\exp\Bigl(-\frac{1}{2}\, s^2\,(t-t_0)\Bigr)\;. \tag{8.60}$$

Insertion of the initial condition $\phi(s,t_0) = \exp(\dot{\imath} s\, w_0)$ completes the calculation of the characteristic function

$$\phi(s,t) \;=\; \exp\Bigl(\dot{\imath} s\, w_0 - \frac{1}{2}\, s^2\,(t-t_0)\Bigr)\;. \tag{8.61}$$

The probability density is now obtained through inversion of Fourier transformation

$$p\,(w,t|\,w_0,t_0) \;=\; \frac{1}{\sqrt{2\pi\,(t-t_0)}}\,\exp\left(-\frac{(w-w_0)^2}{2\,(t-t_0)}\right)\;. \tag{8.62}$$

It is straightforward to identify the density function of the Wiener process as a normal distribution with an expectation value of $E\Bigl(\mathcal{W}(t)\Bigr) = w_0 = \nu$ and variance of $E\Bigl(\bigl(\mathcal{W}(t)-w_0\bigr)^2\Bigr) = t-t_0 = \sigma(t)^2$. Accordingly, an initially sharp

where $\xi(t)$ is a stochastic variable representing the noise term. Rigorous mathematical handling of this equation leads to problems related to the fact that the Wiener process and $\xi(t)$ are nowhere differentiable and integration raises problems that can be avoided by the formulation (8.59).

[26]The characteristic of a probability distribution is defined as

$$\phi(s) \;=\; \int_{-\infty}^{+\infty} \exp(\dot{\imath} s x)\, dF(x) \;=\; \int_{-\infty}^{+\infty} \exp(\dot{\imath} s x)\, f(x)\, \mathrm{d}x\;,$$

and fulfils $\phi(s) = E(e^{\dot{\imath} s \mathcal{X}})$. Apart from constant factors the characteristic function is the Fourier transform of the probability density.

distribution of th form of a δ-function spreads in space as time progresses and eventually in the limit $t \to \infty$ the variance diverges. An illustration is shown in Fig. 8.17.

The Wiener process may be characterized by three important features:

(i) irregularity of sample path,

(ii) non-differentiability of sample path, and

(iii) independence of increment.

Although the mean value $E\big(\mathcal{W}(t)\big)$ is well defined and independent of time, w_0, in the sense of a martingale, the expectation value of the mean square of the random variable $E\big(\mathcal{W}(t)^2\big)$ 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 trajectories in Fig. 8.14). We shall encounter such a situation with finite mean but diverging variance also in biology in the case of multiplication as a birth and death process (subsection 9.1.3): Although the mean is well defined it looses its value in practice when the standard deviation becomes much larger than the expectation value.

Continuity of sample paths of the Wiener process has been demonstrated previously. In order to show that the trajectories of the Wiener process are nowhere differentiable we consider the probability

$$P\left(\left|\frac{\mathcal{W}(t+h)-\mathcal{W}(t)}{h}\right| > k\right) \;=\; 2\int_{k\,h}^{\infty} \mathrm{d}w\,\frac{1}{\sqrt{2\pi h}}\,\exp(-w^2/2h)\;,$$

which can be readily computed from the conditional probability (8.62). In the limit $h \to 0$ the integral becomes $\frac{1}{2}$ and the probability is one. The result implies that, no matter what finite k we choose, $|\big(\mathcal{W}(t+h)-\mathcal{W}(t)\big)/h|$ will almost certainly be greater than this value. In other words, the derivative of the Wiener process will be infinite with probability one and the sample path is not differentiable.

The Wiener process is self-similar and this property is often referred to as *Brownian scaling*. We assume that W_{ct} is a Wiener process with the probability density $f_{W_{ct}}(x) = \exp(-x^2/2ct) \,/\, \sqrt{2\pi t}$ and then the process $V_t = W_{ct}\big/\sqrt{c}$ is another Wiener process for every $c > 0$. In other words, a Wiener process looks the same on all time scales.

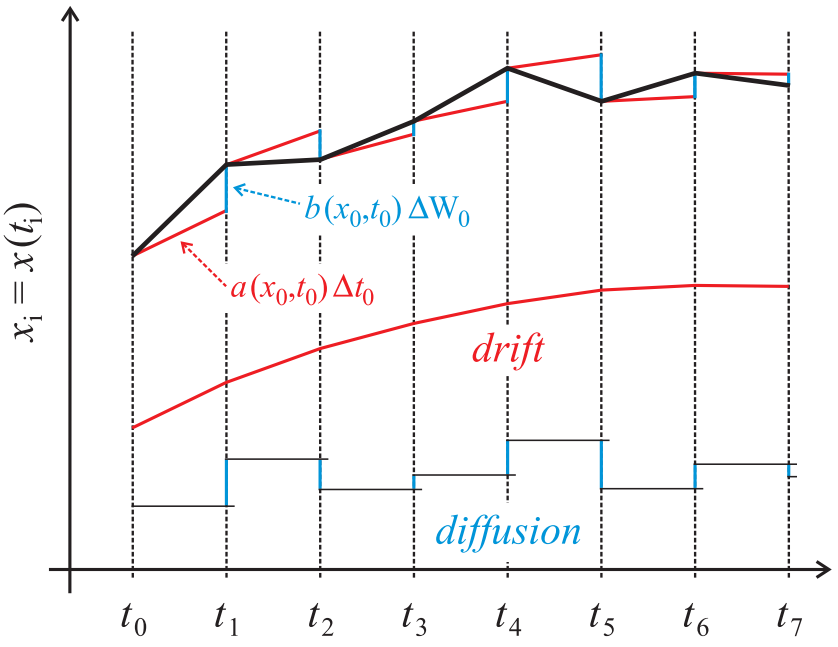


Figure 8.18: **Stochastic integration.** The figure illustrates the Cauchy-Euler procedure for the construction of an approximate solution of the stochastic differential equation (8.63). The stochastic process consists of two different components: (i) the drift term, which is the solution of the ODE 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 ΔW_i are independent or uncorrelated. An approximation to a particular solution of the stochastic process is constructed by letting the mesh size approach zero, $\lim \Delta t \to 0$.

Diffusion is closely related to the Wiener process and hence for the applications in physics and chemistry is important that the increments of $\mathcal{W}(t)$ are statistically independent. The proof makes use of the Markov property of the Wiener process and derives factorizability of increments, which is tantamount to independence of the variables ΔW_i of each other and by the same token of $\mathcal{W}(t_0)$.

Stochastic differential equation. A stochastic variable $x(t)$ is consistent with

an Itō stochastic differential equation (SDE) [159, 160][27]

$$\mathrm{d}x(t) \;=\; a\big(x(t),t\big)\,\mathrm{d}t \;+\; b\big(x(t),t\big)\,\mathrm{d}W(t)\;, \tag{8.63}$$

if for all t and $t_0 = 0$ the integral equation

$$x(t) \;-\; x(0) \;=\; \int_0^t a\big(x(\tau),\tau\big)\,\mathrm{d}\tau \;+\; \int_0^t b\big(x(\tau),\tau\big)\,\mathrm{d}W(\tau) \tag{8.64}$$

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 $x(t_0) = x_0$ by means of a discretized version

$$x_{i+1} \;=\; x_i \;+\; a(x_i,t_i)\,\Delta t_i \;+\; b(x_i,t_i)\,\Delta W_i\;, \tag{8.64'}$$

wherein $x_i = x(t_i)$, $\Delta t_i = t_{i+1} - t_i$, and $\Delta W_i = W(t_{i+1}) - W(t_i)$. Figure 8.18 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 (8.64') and stochastic diffusion component, which is a random 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 in the stochastic term, ΔW_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.

[27] Stochastic integration requires the definition of a reference point. The definition by Itō choosing the reference point at the beginning of the interval is frequently used in the theory of stochastic processes, because it facilitates integration of stochastic differential equations, but Itō calculus is different from conventional calculus. An alternative definition due to the Russian physicist and engineer Ruslan Leontevich Stratonovich [277] and the American mathematician Donald LeRoy Fisk [94] puts the reference point in the middle of the interval and retains thereby the conventional integration formulas, suffers, however, from being more sophisticated.

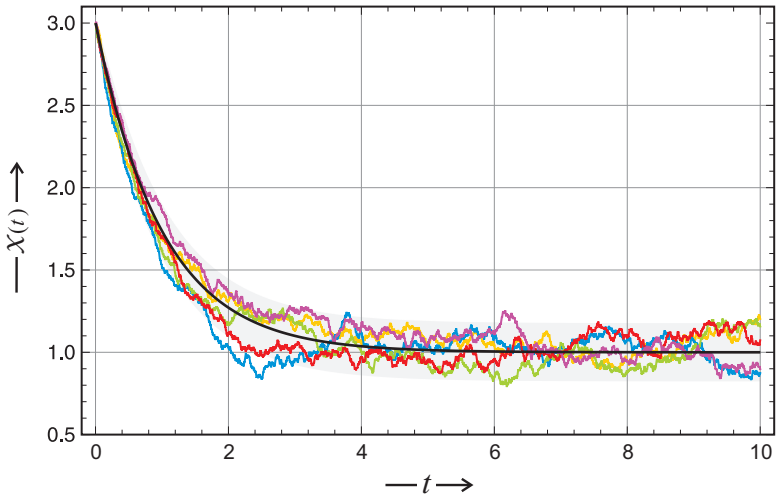


Figure 8.19: **The Ornstein-Uhlenbeck process.** Trajectories are simulated according to $\mathcal{X}_{i+1} = \mathcal{X}_i\, e^{-k\,\vartheta} + \mu(1 - e^{-k\,\vartheta}) + \sigma\sqrt{\frac{1-e^{-2k\,\vartheta}}{2k}}\,(\mathcal{R}_{0,1} - 0.5)$, where $\mathcal{R}_{0,1}$ is a random number drawn by a random number generator from the uniform distribution on the interval $[0, 1]$. The figures shows several trajectories differing only in the choice of seed for *Mersenne Twister* as random number generator,. The black curve represents the expectation value $E\big(\mathcal{X}(t)\big)$ and the area highlighted in grey is the confidence interval $E\pm\sigma$. Choice of parameters: $\mathcal{X}(0) = 3$, $\mu = 1$, $k = 1$, $\sigma = 0.25$, $\vartheta = 0.002$ 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 [116, 290].

Ornstein-Uhlenbeck process. The Ornstein-Uhlenbeck process is named after two Dutch physicists Leonard Ornstein and George Uhlenbeck [288] and represents presumably the simplest stochastic process that approaches a stationary state with a defined variance.[28] The Ornstein-Uhlenbeck process found application also in economics for modeling irregular behavior of financial markets [291]. In essence, the Ornstein-Uhlenbeck process describes exponential relaxation to stationary state or equilibrium superimposed by Brownian motion. Fig. 8.19 presents several trajectories of the Ornstein-Uhlenbeck process, which show nicely the drift and the diffusion component of the in-

[28] The variance of the random walk and the Wiener process diverged in the infinite time limit, $\lim_{t\to\infty} \mathrm{var}\big(\mathcal{W}(t)\big) = \infty$.

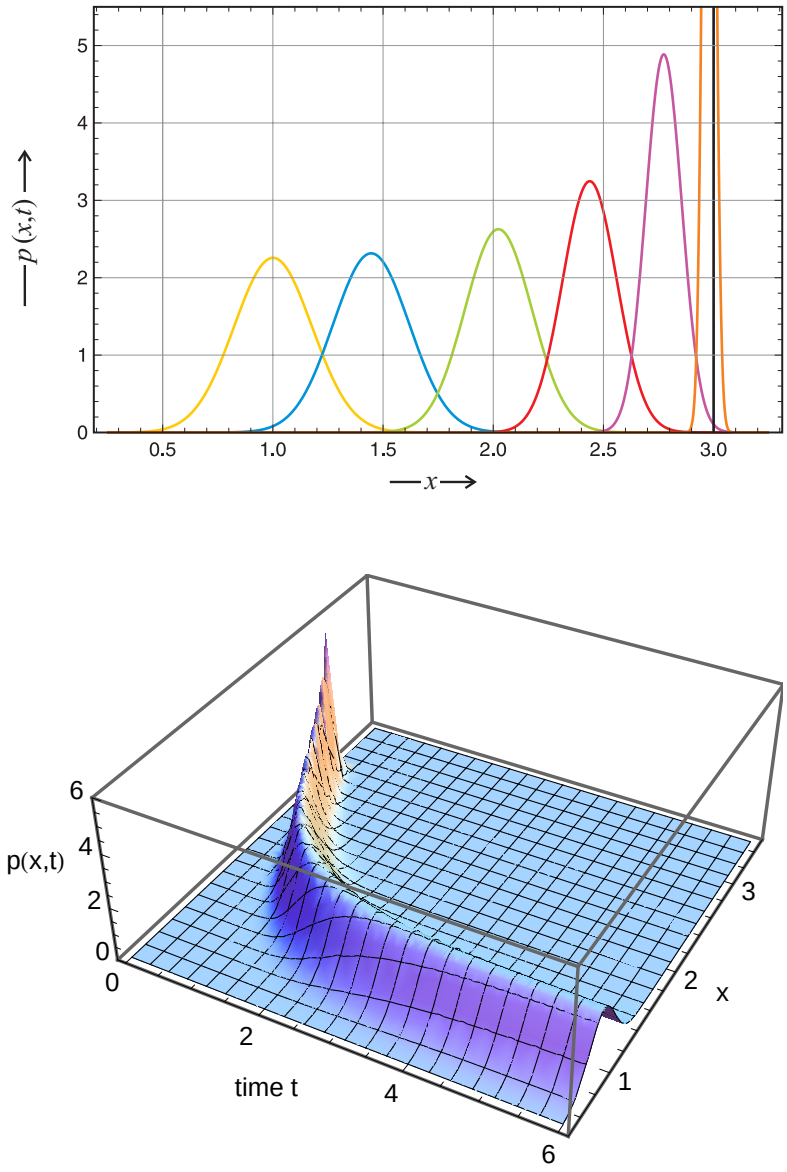


Figure 8.20: **The probability density of the Ornstein-Uhlenbeck process.** Starting from the initial condition $p\,(x,t_0) = \delta(x - x_0)$ (black) the probability density (8.66) broadens and migrates until it reaches the stationary distribution (yellow). The lower plot presents an illustration in 3D. Choice of parameters: $x_0 = 3$, $\mu = 1$, $k = 1$, and $\sigma = 0.25$. Times: $t = 0$ (black), 0.12 (orange), 0.33 (violet), 0.67 (green), 1.5 (blue), and 8 (yellow).

dividual runs.

The Fokker-Planck equation of the Ornstein-Uhlenbeck process for the probability density $p\,(x,t) = p\,(x,t|\,x_0,0)$ is of the form

$$\frac{\partial p\,(x,t)}{\partial t} \;=\; k\,\frac{\partial}{\partial x}\Big((x-\mu)\,p\,(x,t)\Big) \;+\; \frac{\sigma^2}{2}\,\frac{\partial^2 p\,(x,t)}{\partial x^2}\;, \tag{8.65}$$

with k is the rate parameter of the exponential decay, μ the expectation value of the random variable in the long time limit, $\mu = \lim_{t\to\infty} E\big(\mathcal{X}(t)\big)$, and $\sigma^2/(2k)$ being the long time variance. Applying the initial condition $p\,(x,0) = p\,(x,0|\,x_0,0) = \delta(x-x_0)$ the probability density can be obtained by standard techniques

$$p\,(x,t) \;=\; \sqrt{\frac{k}{\pi\sigma^2\,(1-e^{-2k\,t})}}\;\exp\left(-\frac{k}{\sigma^2}\,\frac{(x-\mu-(x_0-\mu)e^{-kt})^2}{1-e^{-2kt}}\right)\;. \tag{8.66}$$

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 indeed recalling a common definition of the Dirac delta-function

$$\delta_\alpha(x) \;=\; \lim_{\alpha\to 0}\,\frac{1}{\alpha\sqrt{\pi}}\;e^{-x^2/\alpha^2}\;, \tag{8.67}$$

and 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:

$$\lim_{t\to\infty} p\,(x,t) \;=\; \sqrt{\frac{k}{\pi\sigma^2}}\,e^{-k\,(x-\mu)^2/\sigma^2}\;,$$

which is a normal density with expectation value μ 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 Fig. 8.20.

The Ornstein-Uhlenbeck process can be modeled efficiently by a stochastic differential equation

$$\mathrm{d}x(t) \;=\; k\,\big(\mu - x(t)\big)\;\mathrm{dt} \;+\; \sigma\,dW(t)\;. \tag{8.68}$$

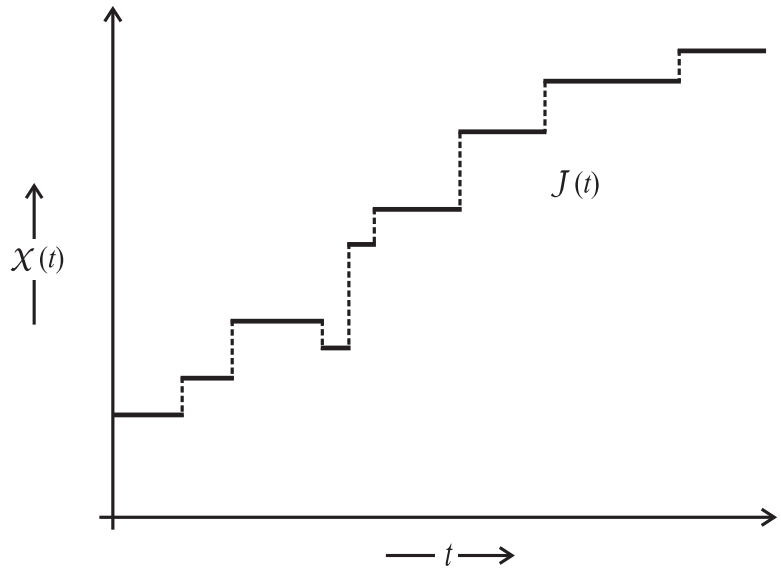


Figure 8.21: **A jump process.** The stochastic variable $\mathcal{X}(t)$ in a jump process $\mathcal{J}(t)$ is defined only for discrete values, integers, for example, representing particle numbers. The step size or height of jumps may be variable in general but in the chemical master equation (see section 9.1) it will be restricted to ± 1 and ± 2 depending on the reaction mechanism. In the pure jump process no continuous changes of stochastic variables are allowed.

The SDE can be used for simulating individual trajectories [116, 290] by means of the following equation

$$\mathcal{X}_{i+1} \,=\, \mathcal{X}_i\, e^{-k\,\vartheta} \,+\, \mu(1-e^{-k\,\vartheta}) \,+\, \sigma\sqrt{\frac{1-e^{-2k\,\vartheta}}{2k}}\,(\mathcal{R}_{0,1}-0.5)\ ,$$

where $\vartheta = \Delta t/n_{\mathrm{st}}$ 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.

Jump process. In a jump process only the last term in the differential Chapman-Kolmogorov Equ. (8.53c) contributes, since we have $\boldsymbol{A}(\boldsymbol{z},t) = 0$ and $\mathrm{B}(\boldsymbol{z},t) = 0$. The resulting equation is known as *master equation*:

$$\frac{\partial}{\partial t}p(\boldsymbol{z},t|\boldsymbol{y},t') \,=\, \int \mathrm{d}\boldsymbol{x}\Big(W(\boldsymbol{z}|\boldsymbol{x},t)\,p(\boldsymbol{x},t|\boldsymbol{y},t') \,-\, W(\boldsymbol{x}|\boldsymbol{z},t)\,p(\boldsymbol{z},t|\boldsymbol{y},t')\Big)\ . \qquad (8.69)$$

In order to illustrate the general process described by the master equation (8.69) we consider the evolution in a short time interval. For this goal we solve approximately to first order in Δt and use the initial condition $p\,(\boldsymbol{z}, t|\,\boldsymbol{y}, t) = \delta(\boldsymbol{y}-\boldsymbol{z})$ representing a sharp probability density at $t = 0$:[29]

$$\begin{aligned} p(\boldsymbol{z}, t+\Delta t|\boldsymbol{y}, t) \;&=\; p(\boldsymbol{z}, t|\boldsymbol{y}, t) + \frac{\partial}{\partial t} p(\boldsymbol{z}, t|\boldsymbol{y}, t)\,\Delta t \;+\ldots\; \approx \\ &\approx\; p(\boldsymbol{z}, t|\boldsymbol{y}, t) + \frac{\partial}{\partial t} p(\boldsymbol{z}, t|\boldsymbol{y}, t)\,\Delta t \;= \\ &= \delta(\boldsymbol{y}-\boldsymbol{z}) + \left(W(\boldsymbol{z}|\boldsymbol{y}, t) - \int \mathrm{d}\boldsymbol{x} W(\boldsymbol{x}|\boldsymbol{y}, t)\,\delta(\boldsymbol{y}-\boldsymbol{z}) \right) \Delta t = \\ &= \left(1 - \int \mathrm{d}\boldsymbol{x} W(\boldsymbol{x}|\boldsymbol{y}, t)\Delta t \right) \delta(\boldsymbol{y}-\boldsymbol{z}) \;+\; W(\boldsymbol{z}|\boldsymbol{y}, t)\,\Delta t\;. \end{aligned}$$

In the first term, the coefficient of $\delta(\boldsymbol{y}-\boldsymbol{z})$ is the (finite) probability for the particle to stay in the original position $\boldsymbol{y}$, whereas the distribution of particles that have jumped is given after normalization by $W(\boldsymbol{z}|\,\boldsymbol{y}, t)$. A typical path $\vec{\mathcal{X}}(t)$ thus will consist of constant sections, $\vec{\mathcal{X}}(t) = \text{const}$, and discontinuous jumps which are distributed according to $W(\boldsymbol{z}|\,\boldsymbol{y}, t)$ (Fig. 8.21). It is worth noticing that a pure jump process does occur here even though the variable $\vec{\mathcal{X}}(t)$ can take on a continuous range of values.

In a special case of the master equations, which nevertheless is fundamental for modeling processes in chemistry and biology, the sample space is mapped onto the space of integers, $\Omega \to \mathbb{Z} = \{.., -2, -1, 0, 1, 2, ..\}$. Then, we can use conditional probabilities rather than probability densities in the master equation:

$$\frac{\partial P(\boldsymbol{n}, t|\boldsymbol{n}', t')}{\partial t} = \sum_m \Big(W(\boldsymbol{n}|\boldsymbol{m}, t)\,P(\boldsymbol{m}, t|\boldsymbol{n}', t') - W(\boldsymbol{m}|\boldsymbol{n}, t)\,P(\boldsymbol{n}, t|\boldsymbol{n}', t') \Big). \quad (8.69')$$

Clearly, this process is confined to jumps since only discrete values of the random variable $\vec{\mathcal{N}}(t)$ are allowed. The master equation on the even more restricted sample space $\Omega \to \mathbb{N}^0 = \{0, 1, 2, \ldots\}$ is of particular importance in chemical kinetics. The random variable $\mathcal{N}(t)$ then counts particle numbers which are necessarily non-negative integers. Furthermore, conservation of

[29] We recall a basic property of the delta-function: $\int f(x)\delta(x-y)\,dx \;=\; f(y)$.

mass introduces an upper bound, for the random variable, $\mathcal{N} \leq N$, and restricting reaction kinetics to elementary step limits the changes in particle numbers to $\Delta\mathcal{N} = 0, \pm 1$, and ± 2.

Noting initial conditions separate from the differential equation, recalling that time t is the only continuous variable when diffusion and other spatial phenomena are not considered, and introducing the physical limitations on particle numbers we obtain eventually

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; \sum_{m=n-2}^{n+2} \Big(W(n|m,t)\, P_m(t \;-\; W(m|n,t)\, P_n(t) \Big) \tag{8.70}$$

The transition probabilities are assumed to be time-dependent here and this is commonly not the case in chemistry and here we shall assume time independence in general. The transition probabilities $W(n|m)$ are understood as the elements of a transition matrix

$$\mathrm{W} \doteq \{W_{nm}; n, m \in \mathbb{N}^0\} \tag{8.71}$$

Diagonal elements W_{nn} cancel in the master Equ. (8.70) and hence need not be defined. According to their nature as *transition probabilities*, all W_{nm} with $n \neq m$ have to be nonnegative. Two definitions of the diagonal elements are common (i) normalization

$$W_{nn} \;=\; 1 \;-\; \sum_{m \neq n} W_{mn} \quad \text{with} \quad \sum_{m} W_{mn} \;=\; 1$$

as used for example in the mutation selection problem (4.9), or (ii) we may define $\sum_m W_{mn} = 0$ which implies $W_{nn} = -\sum_{m \neq n} W_{mn}$ and then insertion into (8.70) leads to a compact form of the master equation

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; \sum_{m} W_{nm} P_m(t) \;. \tag{8.70'}$$

In subsection 9.1.1 we shall discuss specific examples of chemical master equations and methods to drive analytical solutions. Here, we illustrate by means of a simple example that became a standard problem in stochasticity:

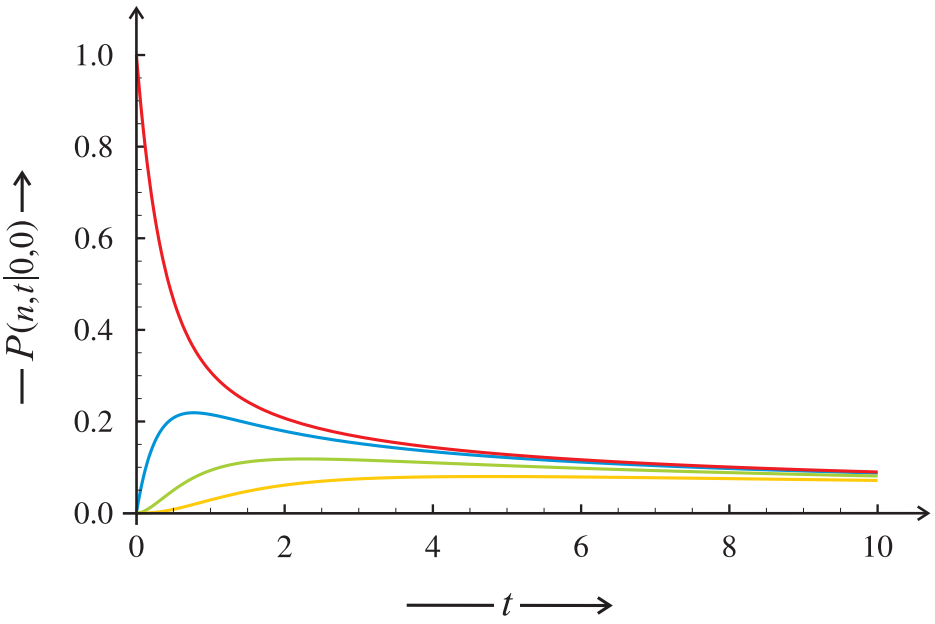


Figure 8.22: **Probability distribution of the random walk.** The figure presents the conditional probabilities $P(n,t|0,0)$ of a random walker to be in position $n \in \mathbb{Z}$ at time t for the initial condition to be at $n = 0$ at time $t = t_0 = 0$. The n-values of the individual curves are: $n = 0$ (black), $n = 1$ (blue), $n = 2$ (purple), and $n = 3$ (red). Parameter choice: $\vartheta = 1$, $l = 1$.

Random walk in one dimension. The random walk in one dimension expressed by the random variable $\mathcal{R}(t)$ describes the movement along a straight line by taking steps with equal probability to the left or to the right. The length of the steps is l and the position of the walker is recorded as a function of time (see Fig. 8.14, where the random walk has been used as illustration for a martingale). The process is discrete in space – the positions that can be reached are integer multiples of the elementary length l: $\mathcal{R}(t) = n\,l$ with n being an integer, $n \in \mathbb{Z}$ and continuous in time t. The first problem we want to solve is the calculation of the probability that the walk reaches the point at distance $n\,l$ at time t when it started at the origin at time $t_0 = 0$. For this goal we cast the random walk in a master equation and search for an analytical solution.

In the master equation we have the following transition probabilities per

unit time provided the time step δt has been chosen sufficiently small, such that only single steps are occurring:

$$W(n+1|n,t) \,=\, W(n-1|n,t) \,=\, \vartheta\,,\; W(m|n,t) \,=\, 0\;\forall\; m \neq \{n+1, n-1\}\;. \quad (8.72)$$

The master equation describing the probability of the walk being in position $n\,l$ at time t when it started at $n'\,l$ at time t' is

$$\frac{\partial P(n,t|n',t')}{\partial t} \;=\; \vartheta\,\Big(P(n+1,t|n',t') + P(n-1,t|n',t') - 2\,P(n,t|n',t')\Big)\;. \quad (8.73)$$

In order to solve the master equation we introduce the time dependent characteristic function:

$$\phi(s,t) \;=\; E(e^{\dot{\imath}s\,n(t)}) \;=\; \sum_n P(n,t|n',t')\,\exp(\dot{\imath}s\,n)\;. \quad (8.74)$$

Combining (8.73) and (8.74) yields

$$\frac{\partial\phi(s,t)}{\partial t} \;=\; \vartheta\,\big(e^{\dot{\imath}s} + e^{-\dot{\imath}s}\big)\,\phi(s,t)\;,$$

and the solution for the initial condition $n' = 0$ at $t' = 0$ takes on the form

$$\phi(s,t) \;=\; \phi(s,0)\,\exp\Big(\vartheta\,t\,(e^{\dot{\imath}s} + e^{-\dot{\imath}s} - 2)\Big) \;=\; \exp\Big(\vartheta\,t\,(e^{\dot{\imath}s} + e^{-\dot{\imath}s} - 2)\Big)\;.$$

Comparison of coefficients for the individual powers of $e^{\dot{\imath}s}$ yields the individual conditional probabilities:

$$\begin{aligned} P(n,t|0,0) \;&=\; I_n(4\vartheta t)\,e^{-2\vartheta t}\,,\; n\in\mathbb{Z} \;\text{ or} \\ P_n(t) \;&=\; I_n(4\vartheta t)\,e^{-2\vartheta t}\,,\; n\in\mathbb{Z} \;\text{ for }\; P_n(0) \,=\, \delta(n)\;. \end{aligned} \quad (8.75)$$

where the pre-exponential term is written in terms of modified Bessel functions $I_k(\theta)$ with $\theta = 4\vartheta t$, which are defined by

$$I_k(\theta) \;=\; \sum_{j=0}^{\infty} \frac{(\theta/4)^{2j+k}}{j!(j+k)!}\;. \quad (8.76)$$

It is straightforward to calculate first and second moments from the characteristic function $\phi(s,t)$:

$$E\big(\mathcal{R}(t)\big) \;=\; n_0 \quad\text{and}\quad \sigma^2\big(\mathcal{R}(t)\big) \;=\; 2\vartheta\,(t-t_0)\;. \quad (8.77)$$

The expectation value is constant and coincides with the starting point of the random walk and the variance increases linearly with time. In Fig. 8.14 we showed individual trajectories of the random walk together with the expectation value and the $\pm\sigma$ confidence interval. In Fig. 8.22 we illustrate the probabilities $P_n(t)$ by means of a concrete example. The probability distribution is symmetric for a symmetric initial condition $P_n(0) = \delta(n)$ 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 entire spatial domain. In case $n \in \mathrm{Z}$ all probabilities vanish: $\lim_{t\to\infty} P_n(t) = 0$ for all n.

It is straightforward to consider the continuous time random walk in the limit of continuous space. This is achieved by setting the distance traveled to $x = n\,l$ and performing the limit $l \to 0$. For that purpose we can start from the characteristic function of the distribution in x,

$$\phi(s,t) \;=\; E\Big(e^{\imath sx}\Big) \;=\; \phi(ls,t) \;=\; \exp\Big(\vartheta\,(e^{\imath ls} + e^{-\imath ls} - 2)\Big)\;,$$

and take the limit of infinitesimally small steps, $\lim l \to 0$:

$$\begin{aligned} &\lim_{l\to 0} \exp\Big(\vartheta\,t\,(e^{\imath ls} + e^{-\imath ls} - 2)\,t\Big) \;= \\ =\; &\lim_{l\to 0} \exp\Big(\vartheta\,t\,(-l^2 s^2 \,+\, \ldots)\Big) \;=\; \exp(-s^2\,Dt/2)\;, \end{aligned}$$

where we used the definition $D = 2\lim_{l\to 0}(l^2\vartheta)$. Since this is the characteristic function of the normal distribution we obtain for the density (8.30):

$$p\,(x,t|\,0,0) \;=\; \frac{1}{\sqrt{2\pi Dt}}\,\exp\big(-x^2/2Dt\big)\;. \qquad (8.30')$$

We could also have proceeded directly from equation (8.73) and expanded the right-hand side as a function of x up to second order in ℓ which gives

$$\frac{\partial p\,(x,t|\,0,0)}{\partial t} \;=\; \frac{D}{2}\,\frac{\partial^2}{\partial x^2}\,p\,(x,t|\,0,0)\;, \qquad (8.78)$$

where D stands again for $2\lim_{l\to 0}(l^2\vartheta)$. This equation is readily recognized as the special Fokker-Planck equation for the diffusion problem, which was be considered in detail in the paragraph on the Wiener process.

8.2.3 Lévy processes

A class of stochastic processes is named after the French mathematician Paul Lévy, which are derived from the differential Chapman-Kolmogorov equation (8.53) by making the assumption of homogeneity in time and probability space. In other words the functions in (8.53) are assumed to be constants. For one dimension we find:

$$A(x,t) \implies a \ , \tag{8.79a}$$

$$B(x,t) \implies \frac{1}{2}\,\sigma^2 \ , \quad \text{and} \tag{8.79b}$$

$$W(z|\,x,t) \implies w(z-x) \ . \tag{8.79c}$$

With these assumptions equations (8.53) becomes[30]

$$\begin{aligned}\frac{\partial p(z,t)}{\partial t} \ &= \ -\ a\,\frac{\partial p(z,t)}{\partial z} \ + \ \frac{1}{2}\,\sigma^2\,\frac{\partial^2 p(z,t)}{\partial z^2} \ + \\ &+ \ -\!\!\!\!\!\!\int \mathrm{d}u\, w(u)\,\Big(p(z-u,t) - p(z,t)\Big) \ ,\end{aligned} \tag{8.80}$$

where $-\!\!\!\!\!\!\int \mathrm{d}u$ denotes the principal value of the integral in the complex plane.

The characteristic function of a Lévy process can be obtained in explicit form:

$$\phi(s,t) \ = \ \int_{-\infty}^{+\infty} \mathrm{d}z\, e^{\dot{\imath}sz}\, p\,(z,t) \ ,$$

which combined with (8.80) yields

$$\frac{\partial \phi(s,t)}{\partial t} \ = \ \Big(\dot{\imath}\,a\,s \ - \ \frac{1}{2}\,\sigma^2\,s^2 \ + \ -\!\!\!\!\!\!\int_{-\infty}^{+\infty} \mathrm{d}u\,\big(e^{\dot{\imath}su} - 1\big)\,w(u)\Big)\,\phi(s,t) \ .$$

Eventually, the characteristic function for the initial condition $p\,(z,0) = \delta(0)$ or $\mathcal{Z}(0) = 0$ takes on the form

$$\begin{aligned}\phi(s,t) \ &= \ \int_{-\infty}^{+\infty} \mathrm{d}z\, e^{\dot{\imath}sz}\, p\,(z|\,0,t) \ = \\ &= \ \exp\!\left(\Big(\dot{\imath}\,a\,s \ - \ \frac{1}{2}\,\sigma^2\,s^2 -\!\!\!\!\!\!\int_{-\infty}^{+\infty} \mathrm{d}u\,\big(e^{\dot{\imath}su} - 1\big)\,w(u)\Big)\,t\right) \ ,\end{aligned} \tag{8.81}$$

[30] The notation $p(z,t)$ is used as a short version of $p(z|\,y,t)$.

which turns out to be quite useful for analyzing special cases.

Lévy processes are or interest because several fundamental processes like the Wiener process and the Poisson process belong to this class. They received plenty of attention outside science and became really important in financial applications like modeling the financial markets, in particular the stock exchange market [107, pp.235-263].

Closing this section on stochastic processes we recall that three different behaviors in the limit of vanishing time intervals resulted from the differential form of the Chapman-Kolmogorov equation and refer to three classes of processes characterized as (i) drift processes, (ii) diffusion processes , and (iii) jump processes. The jump process will turn to be most relevant for our intentions here because it is the basis of the chemical master equation (section 9.1) and it provides the most straightforward access to handling stochasticity in biology. The Fokker-Planck equation will be encountered again in the discussion of Motoo Kimura's neutral theory of evolution.

9. Stochasticity in chemical reactions

Conventional chemical kinetics as said before does not require rigorous stochastic description, because fluctuations are extremely small and ample is fluctuation spectroscopy, which aims at direct recordings of fluctuations [315, 321]. In particular, very accurate measurements of fluorescence correlation signals have been successful [63, 69, 190, 240]. The breakthrough was achieved here through the usage of high-performance lasers. Direct experimental observations of individual processes are now accessible by means of single molecule techniques. The best studied examples are, for example, the experimental recording of single trajectories of biopolymer folding and unfolding (see, for example, [121, 148, 163, 323]).

The role of fluctuations is especially interesting in situations where fluctuations can be amplified by reaction dynamics. Such situations occur with reaction mechanisms involving nonlinear terms and giving rise to highly sensitive oscillations and deterministic chaos. Other examples are pattern formation in space and time, Turing patterns, migrating spirals, and spatiotemporal chaos. Here, we focus on modeling stochastic chemical phenomena by means of master equations, which have the advantage to be accessible to numerical computation. Master equations applied in chemical kinetics are discussed first, we pay particular attention to the Poisson process that is essential for the analysis of probability distributions of independent events, next we discuss master equations derived from birth-and-death processes, then we review exactly solvable cases of elementary steps in chemical kinetics, and eventually we present and discuss an efficient numerical method for the calculation of stochastic trajectories of chemical reaction networks.

9.1 Master equations in chemistry

Chemical reactions are defined by mechanisms, which can be decomposed into elementary processes. An elementary process describes the transformation of one or two molecules into products. Elementary processes involving three or more molecules are unlikely to happen in the vapor phase or in dilute solutions, because trimolecular encounters are very rare under these conditions. Therefore, elementary steps of three molecules are not considered in conventional chemical kinetics.[1] Two additional events which occur in open systems, for example in flow reactors, are the creation of a molecules through influx or the annihilation of a molecule through outflux. Common elementary steps are:

$$\star \longrightarrow \mathbf{A} \tag{9.1a}$$

$$\mathbf{A} \longrightarrow \oslash \tag{9.1b}$$

$$\mathbf{A} \longrightarrow \mathbf{B} \tag{9.1c}$$

$$\mathbf{A} \longrightarrow 2\,\mathbf{B} \tag{9.1d}$$

$$\mathbf{A} \longrightarrow \mathbf{B} + \mathbf{C} \tag{9.1e}$$

$$\mathbf{A} + \mathbf{B} \longrightarrow \mathbf{C} \tag{9.1f}$$

$$\mathbf{A} + \mathbf{B} \longrightarrow 2\,\mathbf{A} \tag{9.1g}$$

$$\mathbf{A} + \mathbf{B} \longrightarrow \mathbf{C} + \mathbf{B} \tag{9.1h}$$

$$\mathbf{A} + \mathbf{B} \longrightarrow \mathbf{C} + \mathbf{D} \tag{9.1i}$$

$$2\,\mathbf{A} \longrightarrow \mathbf{B} \tag{9.1j}$$

$$2\,\mathbf{A} \longrightarrow 2\,\mathbf{B} \tag{9.1k}$$

$$2\,\mathbf{A} \longrightarrow \mathbf{B} + \mathbf{C} \tag{9.1l}$$

Depending on the number of reacting molecules the elementary processes are called *mono*-, *bi*-, or *trimolecular*. Tri- and higher molecular elementary

[1] Exceptions are reactions involving surfaces as third partner, which are important in gas phase kinetics, and biochemical reactions involving macromolecules.

steps are excluded in conventional chemical reaction kinetics as said above.

The example show in Equ. (9.1g) is a simple *autocatalytic* elementary process. In practice autocatalytic chemical reactions commonly involve many elementary steps and thus are the results of complex reaction mechanisms. In evolution reproduction and replication are obligatory autocatalytic processes, which involve a complex molecular machinery. In order to study the basic features of autocatalysis or chemical self-enhancement and self-organization, single step autocatalysis is often used as in model systems. One particular trimolecular autocatalytic process invented by Ilya Prigogine and Gregoire Nicolis

$$2\,\mathbf{A} + \mathbf{B} \longrightarrow 3\,\mathbf{A}\,, \tag{9.2}$$

became very famous [223] despite its trimolecular nature, which makes it unlikely to occur in real systems. The elementary step (9.2) is the essential step in the so-called Brusselator model that can be analyzed straightforwardly by rigorous mathematical techniques. The Brusselator model gives rise to complex dynamical phenomena in space and time, which are rarely observed in standard 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 named after the English computer scientist Alan Turing [287] and is frequently used as a model for certain classes of pattern formation in biology [204]. The original concept Turing's was based on a reaction-diffusion equation and aimed at a universal model of morphogenesis in development. It has been worked out later in great detail by Alfred Gierer, Hans Meinhardt and others [110, 205]. Molecular genetics, however, has shown that nature is using direct genetic control through cascades of gene regulation rather than self-organized reaction-diffusion patterns [111, 192].

9.1.1 Chemical master equations

Provided particle numbers are used as variables to model the progress of chemical reactions, stochastic variables $\mathcal{N}(t)$ are described by the probabil-

ities $P_n(t) = \text{Prob}\,(\mathcal{N}(t) = n)$ and take only nonnegative integer values, $n \in \mathbb{N}^0$. As mentioned already in the previous subsection 8.2.2 we introduce a few simplifications and conventions in th notation: (i) We shall use the forward equation unless stated differently, (ii) we assume an infinitely sharp initial density, $P(n,0|\,n_0,0) = \delta_{n,n_0}$ with $n_0 = n(0)$, and (iii) we simplify the full notation $P(n,t|\,n_0,0) \;\Rightarrow\; P_n(t)$ with the implicit assumption of the sharp initial condition (ii). Handling extended probability densities as initial conditions will be discussed explicitly. The expectation value of the stochastic variable $\mathcal{N}(t)$ is denoted by

$$E\big(\mathcal{N}(t)\big) = \langle n(t)\rangle = \sum_{n=0}^{\infty} n \cdot P_n(t)\ , \tag{9.3}$$

and its stationary value, provided it exists, will be written

$$\bar{n} = \lim_{t\to\infty} \langle n(t)\rangle\ . \tag{9.4}$$

Almost always $\bar{n}$ will be identical with the long time value of the corresponding deterministic variable. The running index of integers will be denoted by either m or n'. Then the *chemical master equation* is of the previously presented form

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; \sum_m \Big(W(n|m)P_m(t) - W(m|n)P_n(t)\Big) \;=\; \sum_m W_{nm}P_m(t)\,, \tag{8.70}$$

where the compact form requires the definition $\sum_m W_{mn} = 0$ and accordingly $W_{nn} = -\sum_{m\neq n} W_{mn}$.

Introducing vector notation, $\boldsymbol{P}(t) = (P_1(t), \ldots, P_n(t), \ldots)$, we obtain

$$\frac{\partial \boldsymbol{P}(t)}{\partial t} \;=\; \mathrm{W} \times \boldsymbol{P}(t)\ . \tag{8.70"}$$

With the initial condition $P_n(0) = \delta_{n,n_0}$ we can formally solve Equ. (8.70") for each n_0 and obtain

$$P(n,t|\,n_0,0) \;=\; \big(\exp(\mathrm{W}\,t)\big)_{n,n_0}\ ,$$

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 evaluation of this equation boils down to diagonalize the matrix W which can be done analytically in rather few cases only.

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 [115]. The two general requirements that have to be fulfilled are: (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 he developed an algorithm that allows for the computation of trajectories, converges for statistics of trajectories to the exact solution and can be easily implemented for computer simulation. We shall discuss the Gillespie formalism together with the computer program in section 9.3. Here we present analytical solutions of master equations by means of a few selected examples (subsection 9.2).

9.1.2 The Poisson process

A *Poisson process* is a stochastic process that counts the number of events of a certain class by means of a stochastic variable $\mathcal{X}(t)$ and the times at which these events occur. Examples of Poisson processes is the radioactive decay, the arrival of telephone calls, and other discrete events they can be assumed to occur independently. A homogeneous Poisson process is characterized by its rate parameter λ that represents the expected number of arrivals or events per unit time.[2] The probability density of the random variable of a Poisson process expressed as the number of events in the time interval $]t, t+\tau]$, k, is given by a Poisson distribution with the parameter $\lambda\tau$:

$$P\Big(\big(\mathcal{X}(t+\tau) - \mathcal{X}(t)\big) = k\Big) \;=\; \frac{e^{-\lambda\tau}\,(\lambda\tau)^k}{k!} \;\text{ with }\; k \in \mathbb{N}^0 \;. \tag{9.5}$$

The Poisson process can be understood as the simplest example of a Lévy process with $a = \sigma = 0$, and $w(u) = \lambda\,\delta(u-1)$ or as a pure-birth process

[2] In the non-homogeneous Poisson process the rate function depends on time $\lambda(t)$ and the expected number of events in the time interval between a and b is $\lambda_{a,b} = \int_a^b \lambda(t)\,\mathrm{dt}$.

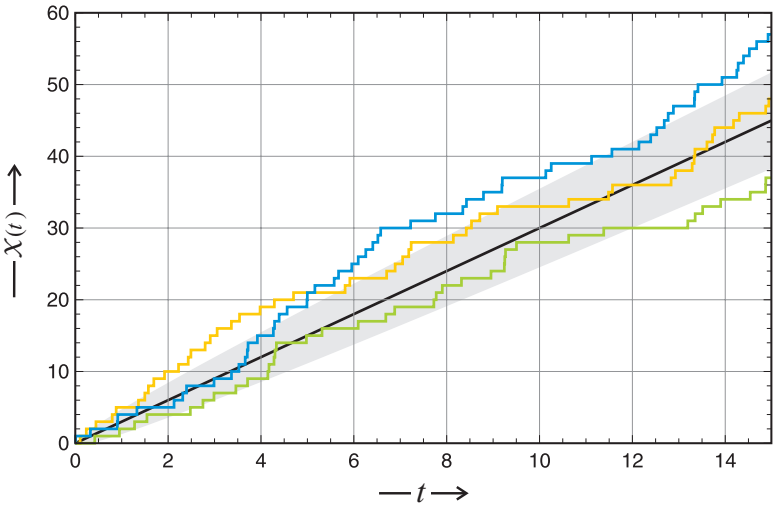


Figure 9.1: **The Poisson process.** The random variable $\mathcal{X}(t)$ of a Poisson process describes consecutive times of occurrence for a series of independent events. The step size of change $\Delta\mathcal{X} = +1$ since multiple occurrence of events at the same instant is excluded. The three trajectories in the figure were obtained by drawing exponentially distributed random real numbers. The expectation value (black) is $E(\mathcal{X}(t) = \lambda t$ and the confidence interval $E \pm \sigma$ with $\sigma = \sqrt{\lambda t}$ is the grey shaded zone. Choice of parameters: $\lambda = 3$. Random seeds ("Mersenne Twister"): 491 (yellow), 733 (blue), and 919 (green).

– this is a birth-and-death process with birth rate λ and zero death rate (subsection 9.1.3). Sometimes the condition of a normalizable function $w(u)$ is stressed by calling the process a *compound Poisson process* because it fulfils $\int_{-\infty}^{+\infty} w(u)\,\mathrm{d}u \equiv \lambda < \infty$. The quantity λ is also called the *intensity* of the process, and it is equal to the inverse mean time between two events.

Because of its general importance we present a precise definition of the Poisson process as a family of random variables $\mathcal{X}(t)$ with t being a continuous variable over the domain $[0, \infty[$ with a parameter λ iff it satisfies the following conditions [36, pp.199-210]

(i) $\mathcal{X}(0) = 0$,

(ii) the increments $\mathcal{X}(t_i+\tau_i)-\mathcal{X}(\tau_i)$ over an arbitrary set of disjoint intervals $(t_i, t_i + \tau_i)$ are independent random variables,

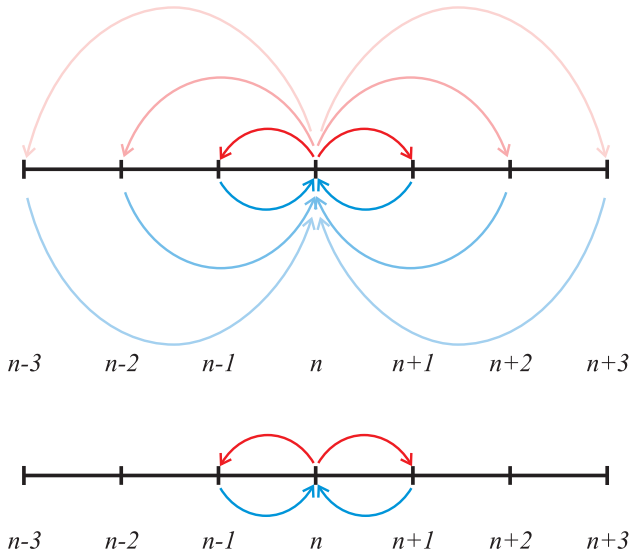


Figure 9.2: **Sketch of the transition probabilities in master equations.** In the general jump process steps of any size are admitted (upper drawing) whereas in birth-and-death processes all jumps have the same size. The simplest and most common case is dealing with the condition that the particles *are born* and *die* one at a time (lower drawing).

(iii) the general increment $\mathcal{X}(t+\tau) - \mathcal{X}(\tau)$ is distributed according to the Poisson density $\text{Pois}(\lambda t)$ for each pair $(t \geq 0, \tau \geq 0)$.

The Poisson process is an excellent model for chemical reactions in a given reaction volume. If we choose the occurrence of reactive collision as event, the individual collisions – no matter whether taking place in the vapor phase or in solution – are to a very good approximation independent events, and the prerequisites for a Poisson process are fulfilled.

9.1.3 Birth-and-death processes

The concept of birth-and-death processes has been created in biology and is based on the assumption that only a finite number of individuals are produced – born – or destroyed – die – in a single event. The simplest and the only case, we shall discuss here, is occurs when birth and death is confined to single

individuals of only one species. These processes are commonly denoted as one step birth-and-death processes.[3] In Fig. 9.2 the transitions in a general jump process and a birth-and-death process are illustrated. Restriction to single events is tantamount to the choice of a sufficiently small time interval of recording, Δt, such that the simultaneous occurrence of two events has a probability of measure zero (see also section 9.3). This small time step is often called the *blind interval*, because no information on things happening within Δt is available.

Then, the transition probabilities can be written in the form

$$W(n|m,t) \;=\; w^+(m)\,\delta_{n,m+1} \;+\; w^-(m)\,\delta_{n,m-1}\;, \tag{9.6}$$

since we are dealing with only two allowed processes:

$n \longrightarrow n+1$ with $w^+(n)$ as transition probability per unit time and

$n \longrightarrow n-1$ with $w^-(n)$ as transition probability per unit time.

Modeling of chemical reactions by birth-and-death processes turns out to be a very useful approach for reaction mechanisms that can be described by changes in a single variable. Another special case of a birth-and-death process is the Poisson process on $n \in \mathbb{N}^0$: It has zero death rate and describes the occurrence of independent events (see also the Poisson distribution in subsection 8.1.3 and the Poisson process in subsection 9.1.2). We shall make use of the Poisson process in describing the occurrence of chemical reaction in subsection 9.3.

The stochastic process can now be described by a birth-and-death master equation

$$\begin{aligned}\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;&= w^+(n-1)\,P_{n-1}(t) \;+\; w^-(n+1)\,P_{n+1}(t) - \\ &\quad - \Big(\big(w^+(n)+w^-(n)\big)\,P_n(t)\;.\end{aligned} \tag{9.7}$$

[3] In addition, one commonly distinguishes between birth-and-death processes in one variable and in many variables [107]. We shall restrict the analysis here to the simpler single variable case here.

There is no general technique that allows to find the time-dependent solutions of equation (9.7) and therefore we shall present some special cases later on. Only few single step birth-and-death processes can be solved analytically.

The stationary case, however, can be analyzed in full generality. Provided a stationary solution of equation (9.7) exists, $\lim_{t\to\infty} P_n(t) = \bar{P}_n$, we can compute it in straightforward manner. It is useful to define a probability current $J(n)$ for the n-th step in the series,

Particle number	0	$\rightleftharpoons$	1	$\rightleftharpoons$	$\ldots$	$\rightleftharpoons$	$n-1$	$\rightleftharpoons$	n	$\rightleftharpoons$	$n+1$ $\ldots$
Reaction step		1		2	$\ldots$	$n-1$		n		$n+1$	$\ldots$

which is of the form

$$J(n) \;=\; w^{-}\, n\, \bar{P}(n) \;-\; w^{+}\,(n-1)\,\bar{P}(n-1)\;. \tag{9.8}$$

The conditions for the stationary solution are given by vanishing time derivatives of the probabilities

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} = \;0\; =\; J(n+1) \;-\; J(n)\;. \tag{9.9}$$

Restriction to positive particle numbers, $n \in \mathbb{N}^0$, implies $w^{-}(0) = 0$ and $P_n(t) = 0$ for $n < 0$, which in turn leads to $J(0) = 0$. Now we sum the vanishing flow terms according to equation (9.9) and obtain:

$$0 \;=\; \sum_{z=0}^{n-1}\bigl(J(z+1) \;-\; J(z)\bigr) \;=\; J(n) \;-\; J(0)\;.$$

Accordingly we find $J(n) = 0$ for arbitrary n, which leads to

$$\bar{P}_n \;=\; \frac{w^{+}(n-1)}{w^{-}(n)}\,\bar{P}_{n-1} \quad \text{and finally} \quad \bar{P}_n \;=\; \bar{P}_0 \prod_{z=1}^{n} \frac{w^{+}(z-1)}{w^{-}(z)}\;.$$

The condition $J(n) = 0$ for every reaction step is known in chemical kinetics as the *principle of detailed balance*: At equilibrium the reaction flow has to vanish for every reaction step. The principle of detailed balance was formulated first by the American mathematical physicist Richard Tolman [286] (see also [107, pp.142-158]).

The macroscopic rate equations are readily derived from the master equation through calculation of the expectation value:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{dt}}\,E\big(n(t)\big) \;&= \frac{\mathrm{d}}{\mathrm{dt}}\left(\sum_{n=0}^{\infty} n\,P_n(t)\right) \;= \\
&= \sum_{n=0}^{\infty} n\left(w^+(n-1)\,P_{n-1}(t) \;-\; w^+(n)\,P_n(t)\right) \;+ \\
&\quad + \sum_{n=0}^{\infty} n\left(w^-(n+1)\,P_{n+1}(t) \;-\; w^-(n)\,P_n(t)\right) \;= \\
&= \sum_{n=0}^{\infty}\Big((n+1)\,w^+(n) \;-\; n\,w^+(n) \;+\; (n-1)\,w^-(n) \;-\; n\,w^-(n)\Big)\,P_n(t) \;= \\
&= \sum_{n=0}^{\infty} w^+(n)\,P_n(t) \;-\; \sum_{n=0}^{\infty} w^-(n)\,P_n(t) \;=\; E\big(w^+(n)\big) \;-\; E\big(w^-(n)\big)\;.
\end{aligned}$$

Neglect of fluctuations yields the deterministic rate equation of the birth-and-death process

$$\frac{\mathrm{d}\langle n\rangle}{\mathrm{dt}} \;=\; w^+\big(\langle n\rangle\big) \;-\; w^-\big(\langle n\rangle\big)\;. \tag{9.10}$$

The condition of stationary yields: $\bar{n} = \lim_{t\to\infty}\langle n(t)\rangle$ for which holds $w^+(\bar{n}) = w^-(\bar{n})$. Compared to this results we note that the maximum value of the stationary probability density, $\max\{\bar{P}_n,\, n\in\mathbb{N}^0\}$, is defined by $\bar{P}_{n+1} - \bar{P}_n \approx -(\bar{P}_n - \bar{P}_{n-1})$ or $\bar{P}_{n+1} \approx \bar{P}_{n-1}$, which coincides with the deterministic value for large n.

9.2 Examples of solvable master equations

Although only a small minority of master equations in chemical reaction kinetics can be solved analytically, these cases provide the best insight into the role of stochasticity in chemical reactions, and they reflect well the repertoire of methodological tricks that are applied in modeling of stochastic processes in chemistry and biology. Here we shall solve the equilibration of the flow in the flow reactor, examples of unimolecular reactions, and one solvable bimolecular reaction.

The flow reactor. In section 4.2 the flow reactor has been introduced as a device for the experimental and theoretical analysis of systems under controlled conditions away from thermodynamic equilibrium (Fig. 4.3). Here we shall use it for analyzing the simplest conceivable process, the non-reactive flow of a single compound, **A**, through the reactor. The stock solution contains **A** at the concentration $[\mathbf{A}]_{\text{influx}} = \hat{a} = \bar{a}$. The influx concentration $\hat{a}$ is equal to the stationary concentration $\bar{a}$, because no reaction is assumed and after sufficiently long time the content of the reactor will be stock solution. The intensity of the flow is measured by means of the flow rate r and this implies an influx of $\hat{a} \cdot r$ of **A** into the reactor, instantaneous mixing with the content of the reactor, and an outflux of the mixture in the reactor at the same flow rate r.[4] If the volume of the reactor is V, the mean residence time of a volume element dV in the reactor is $\tau_R = V \cdot r^{-1}$.

In- and outflux of compound **A** into and from the reactor are modeled by two formal elementary steps or pseudo-reactions

$$\begin{aligned} \star &\longrightarrow \mathbf{A} \\ \mathbf{A} &\longrightarrow \oslash \ . \end{aligned} \tag{9.11}$$

In chemical kinetics the differential equations are almost always formulated in molecular concentrations. For the stochastic treatment, however, we replace concentrations by the numbers of particles, $\bar{n} = \bar{a} \cdot V \cdot N_L$ with $n \in \mathbb{N}^0$ and N_L being Loschmidt's or Avogadro's number,[5] the number of particles per mole.

The particle number of **A** in the reactor is a stochastic variable with the probability $P_n(t) = P\big(\mathcal{N}(t) = n\big)$. The time derivative of the probability

[4] The assumption of equal influx and outflux rate is required because we are dealing with a flow reactor of constant volume V (CSTR, Fig. 4.3).

[5] As a matter of fact there is a difference between Loschmidt's and Avogadro's number that is often ignored in the literature: Avogadro's number, $N_L = 6.02214179 \times 10^{23}\,\text{mole}^{-1}$ refers to one mole substance whereas Loschmidt's constant $n_0 = 2.6867774 \times 10^{25}\,\text{m}^{-3}$ 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 $22.414\ \text{dm}^{-3} \cdot \text{mole}^{-1}$.

distribution is described by means of the master equation

$$\frac{\mathrm{d}P_n(t)}{\mathrm{dt}} \;=\; r\Big(\bar{n}\,P_{n-1}(t) + (n+1)\,P_{n+1}(t) - (\bar{n}+n)\,P_n(t)\Big)\,;\; n \in \mathrm{N}^0\;. \quad (9.12)$$

Here we have implicitly assumed Equ. (9.12) describes a birth-and-death process with $w^+(n) = r\bar{n}$ and $w^-(n) = rn$. Thus we have a constant birth rate and a death rate which is proportional to n. see next subsection 9.1.3

Solutions of the master equation can be found in text books listing stochastic processes with known solutions, for example [118]. Here we shall derive the solution by means of probability generating functions in order to illustrate this particularly powerful approach. The probability generating function of a nonnegative valued random variable is defined by

$$g(s,t) \;=\; \sum_{n=0}^{\infty} P_n(t)\,s^n\;. \quad (9.13)$$

In general the auxiliary variable s is real valued, $s \in \mathbb{R}^1$, although generating functions with complex s can be of advantage. Sometimes the initial state is included in the notation: $g_{n_0}(s,t)$ implies $P_n(0) = \delta_{n,n_0}$. Partial derivatives with respect to time t and the variable s are readily computed:

$$\begin{aligned}
\frac{\partial g(s,t)}{\partial t} \;&=\; \sum_{n=0}^{\infty} \frac{\partial P_n(t)}{\partial t}\cdot s^n \;= \\
&=\; r\sum_{n=0}^{\infty}\Big(\bar{n}\,P_{n-1}(t) \;+\; (n+1)\,P_{n+1}(t) \;-\; (\bar{n}+n)\,P_n(t)\Big)\,s^n \quad \text{and} \\
\frac{\partial g(s,t)}{\partial s} \;&=\; \sum_{n=0}^{\infty} n\,P_n(t)\,s^{n-1}\;.
\end{aligned}$$

Proper collection of terms and arrangement of summations – by taking into account: $w^-(0) = 0$ – yields

$$\frac{\partial g(s,t)}{\partial t} \;=\; r\bar{n}\sum_{n=0}^{\infty}\Big(P_{n-1}(t) \;-\; P_n(t)\Big)\,s^n \;+\; r\sum_{n=0}^{\infty}\Big((n+1)\,P_{n+1}(t) \;-\; n\,P_n(t)\Big)\,s^n\,.$$

Evaluation of the four infinite sums

$$\sum_{n=0}^{\infty} P_{n-1}(t)\, s^n \;=\; s \sum_{n=0}^{\infty} P_{n-1}(t)\, s^{n-1} \;=\; s\, g(s,t)\,,$$
$$\sum_{n=0}^{\infty} P_n(t)\, s^n \;=\; g(s,t)\,,$$
$$\sum_{n=0}^{\infty} (n+1)\, P_{n+1}(t)\, s^n \;=\; \frac{\partial g(s,t)}{\partial t}\,, \quad \text{and}$$
$$\sum_{n=0}^{\infty} n\, P_n(t)\, s^n \;=\; s \sum_{n=0}^{\infty} n\, P_n(t)\, s^{n-1} \;=\; s\, \frac{\partial g(s,t)}{\partial t}\;,$$

and regrouping of terms yields a linear partial differential equation of first order

$$\frac{\partial g(s,t)}{\partial t} \;=\; r\Big(\bar{n}(s-1)\, g(s,t) \;-\; (s-1)\, \frac{\partial g(s,t)}{\partial s}\Big)\;. \tag{9.14}$$

The partial differential equation (PDE) is solved through consecutive substitutions

$$\phi(s,t) \;=\; g(s,t)\, \exp(-\bar{n}\, s) \quad \longrightarrow \quad \frac{\partial \phi(s,t)}{\partial t} \;=\; -r(s-1)\, \frac{\partial \phi(s,t)}{\partial s}\;,$$
$$s-1 \;=\; e^{\rho} \text{ and } \psi(\rho,t) \;=\; \phi(s,t) \quad \longrightarrow \quad \frac{\partial \psi(\rho,t)}{\partial t} \;+\; r\, \frac{\partial \psi(\rho,t)}{\partial \rho}\;.$$

Computation of the characteristic manifold is equivalent to solving the ordinary differential equation (ODE) $r\,\mathrm{d}t \;=\; -\mathrm{d}\rho$. We find: $rt - \rho \;=\; C$ where C is the integration constant. The general solution of the PDE is an arbitrary function of the combined variable $rt - \rho$:

$$\psi(\rho,t) \;=\; f\big(\exp(-rt+\rho)\big) \cdot e^{-\bar{n}} \quad \text{and} \quad \phi(s,t) \;=\; f\Big((s-1)\, e^{-rt}\Big) \cdot e^{-\bar{n}}\;,$$

and the probability generating function

$$g(s,t) \;=\; f\Big((s-1)\, e^{-rt})\Big) \cdot \exp\big((s-1)\bar{n}\big)\;.$$

Normalization of probabilities (for $s \;=\; 1$) requires $g(1,t) \;=\; 1$ and hence $f(0) = 1$. The initial conditions as expressed by the conditional probability $P(n,0|n_0,0) = P_n(0) = \delta_{n,n_0}$ leads to the final expression

$$\begin{aligned} g(s,0) \;&= f(s-1) \cdot \exp\big((s-1)\bar{n}\big) \;=\; s^{n_0}\;, \\ f(\zeta) \;&=\; (\zeta+1)^{n_0} \cdot \exp(-\zeta\bar{n}) \;\text{ with }\; \zeta = (s-1)\, e^{-rt}\;, \\ g(s,t) \;&= \Big(1 + (s-1)\, e^{-rt}\Big)^{n_0} \cdot \exp\big(-\bar{n}(s-1)\, e^{-rt}\big) \cdot \exp\big(\bar{n}(s-1)\big) \;= \\ &= \Big(1 + (s-1)\, e^{-rt}\Big)^{n_0} \cdot \exp\big\{-\bar{n}(s-1)\,(1-e^{-rt})\big\}\;. \end{aligned} \tag{9.15}$$

From the generating function we compute with somewhat tedious but straightforward algebra the probability distribution

$$P_n(t) \;=\; \sum_{k=0}^{\min\{n_0,n\}} \binom{n_0}{k} \bar{n}^{n-k} \cdot \frac{e^{-krt}\left(1-e^{-rt}\right)^{n_0+n-2k}}{(n-k)!} \cdot e^{-\bar{n}\left(1-e^{-rt}\right)} \qquad (9.16)$$

with $n, n_0, \bar{n} \in \mathbb{N}^0$. In the limit $t \to \infty$ we obtain a non-vanishing contribution to the stationary probability only from the first term, $k = 0$, and find

$$\lim_{t\to\infty} P_n(t) \;=\; \frac{\bar{n}^n}{n!} \exp(-\bar{n}) \; .$$

This is a Poissonian distribution with parameter and expectation value $\alpha = \bar{n}$. The Poissonian distribution has also a variance which is numerically identical with the expectation value, $\sigma^2(\mathcal{N}_A) = E(\mathcal{N}_A) = \bar{n}$, and thus the distribution of particle numbers fulfils the $\sqrt{N}$-law at the stationary state.

The time dependent probability distribution allows to compute the expectation value and the variance of the particle number as a function of time

$$\begin{aligned} E\big(\mathcal{N}(t)\big) \;&=\; \bar{n} \,+\, (n_0 - \bar{n}) \cdot e^{-rt} \; , \\ \sigma^2\big(\mathcal{N}(t)\big) \;&=\; \big(\bar{n} \,+\, n_0 \cdot e^{-rt}\big) \cdot \big(1 \,-\, e^{-rt}\big) \; . \end{aligned} \qquad (9.17)$$

As expected the expectation value apparently coincides with the solution curve of the deterministic differential equation

$$\frac{\mathrm{d}n}{\mathrm{dt}} \;=\; w^+(n) \,-\, w^-(n) \;=\; r\,(\bar{n} - n) \; , \qquad (9.18)$$

which is of the form

$$n(t) \;=\; \bar{n} \,+\, (n_0 - \bar{n}) \cdot e^{-rt} \; . \qquad (9.18')$$

Since we start from sharp initial densities variance and standard deviation are zero at time $t = 0$. The qualitative time dependence of $\sigma^2\big(\mathcal{N}_A(t)\big)$, however, depends on the sign of $(n_0 - \bar{n})$:

(i) For $n_0 \le \bar{n}$ the standard deviation increases monotonously until it reaches the value $\sqrt{\bar{n}}$ in the limit $t \to \infty$, and

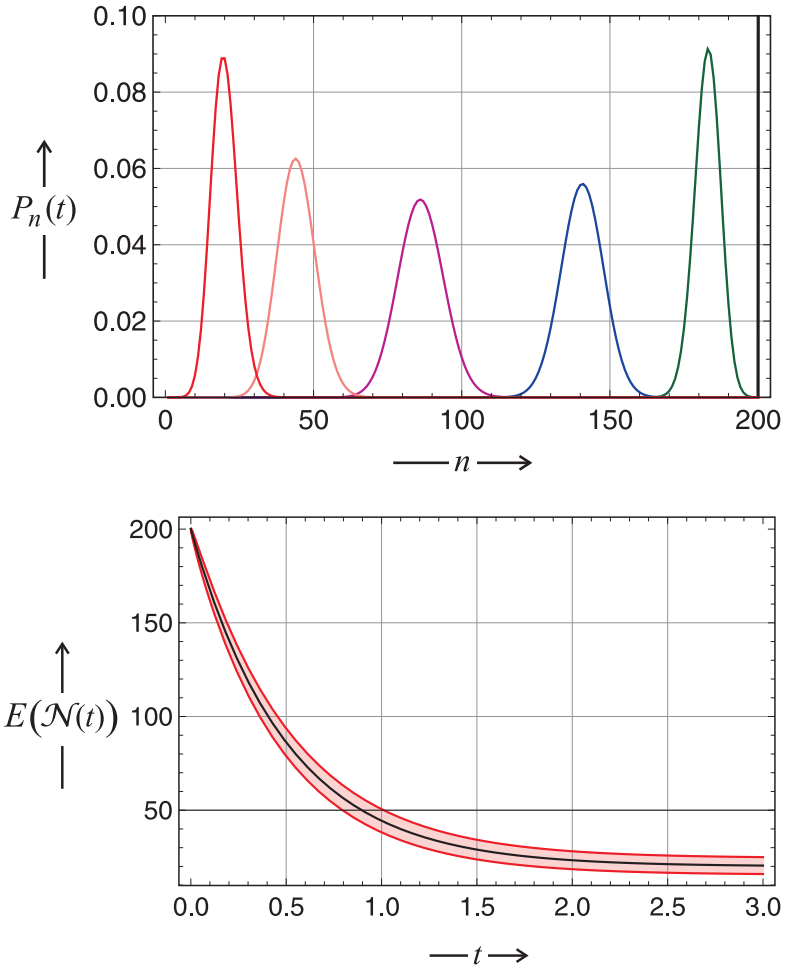


Figure 9.3: **Establishment of the flow equilibrium in the CSTR.** The upper part shows the evolution of the probability density, $P_n(t)$, of the number of molecules of a compound **A** which flows through a reactor of the type illustrated in Fig. 4.3. The initially infinitely sharp density becomes broader with time until the variance reaches its maximum and then sharpens again until it reaches stationarity. The stationary density is a Poissonian distribution with expectation value and variance, $E(\mathcal{N}) = \sigma^2(\mathcal{N}) = \bar{n}$. In the lower part we show the expectation value $E\big(\mathcal{N}(t)\big)$ in the confidence interval $E \pm \sigma$. Parameters used: $\bar{n} = 20$, $n_0 = 200$, 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).

(ii) for $n_0 > \bar{n}$ the standard deviation increases until it passes through a maximum at

$$t(\sigma_{\max}) \;=\; \frac{1}{r}\Big(\ln 2 \,+\, \ln n_0 \,-\, \ln(n_0 - \bar{n})\Big)$$

and approaches the long-time value $\sqrt{\bar{n}}$ from above.

In figure 9.3 we show an example for the evolution of the probability density (9.16). In addition, the figure contains a plot of the expectation value $E\big(\mathcal{N}(t)\big)$ 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 as derived in case of the normal distribution.

The monomolecular reaction. The reversible mono- or unimolecular chemical reaction can be split into two irreversible elementary reactions

$$\mathbf{A} \quad \xrightarrow{k_1} \quad \mathbf{B} \tag{9.19a}$$

$$\mathbf{A} \quad \xleftarrow{k_2} \quad \mathbf{B}\,, \tag{9.19b}$$

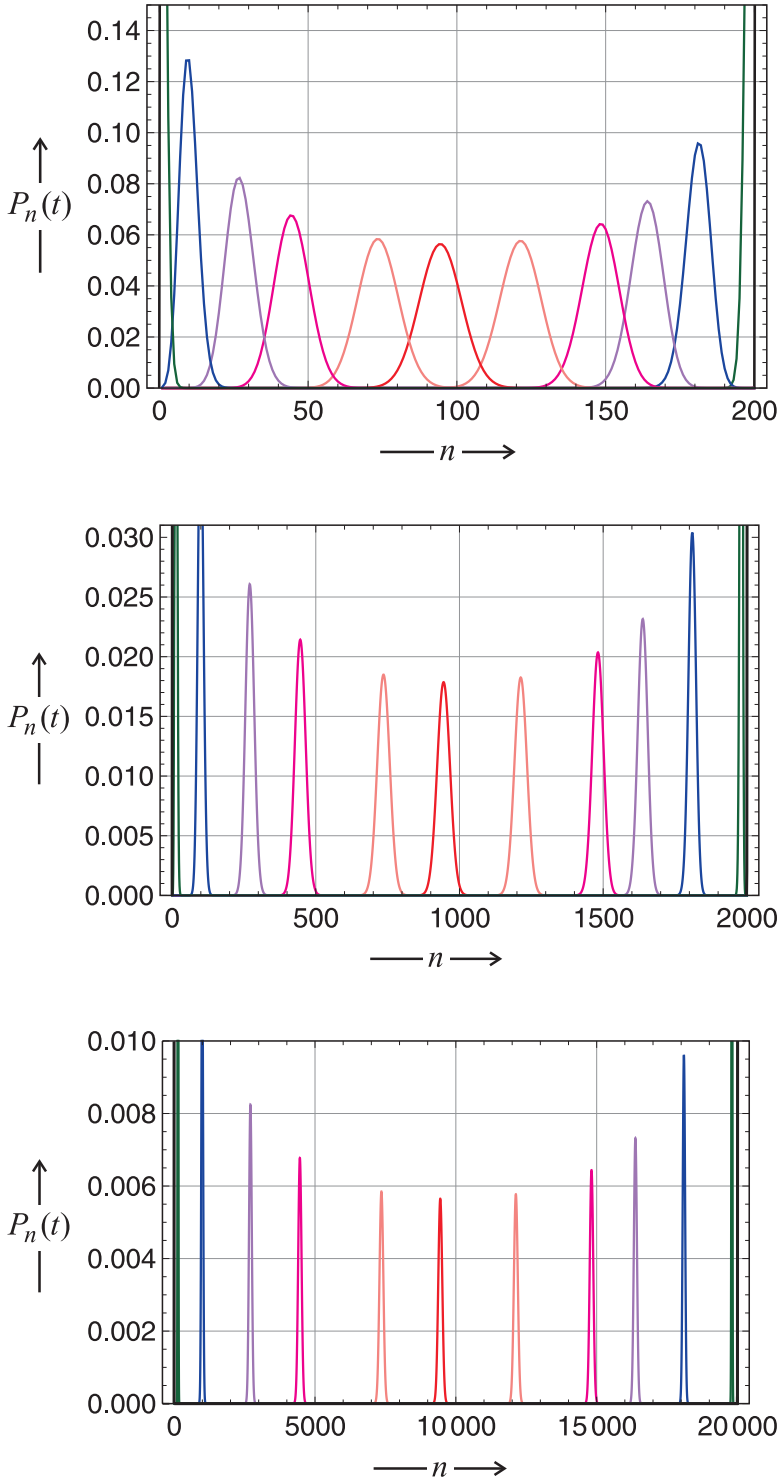
wherein the reaction rate parameters, k_1 and k_2, are called *reaction rate constants*. The reaction rate parameters depend on temperature, pressure, and other environmental factors. At equilibrium the rate of the forward reaction (9.19a) is precisely compensated by the rate of the reverse reaction (9.19b), $k_1 \cdot [\mathbf{A}] = k_2 \cdot [\mathbf{B}]$, leading to the condition for the thermodynamic equilibrium:

$$K \;=\; \frac{k_1}{k_2} \;=\; \frac{[\mathbf{B}]}{[\mathbf{A}]}\,. \tag{9.20}$$

The parameter K is called the *equilibrium constant* that 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:

$$\frac{\mathcal{N}_A(t) \,+\, \mathcal{N}_B(t)}{V \cdot N_A} \;=\; [\mathbf{A}] \,+\, [\mathbf{B}] \;=\; c(t) \;=\; c_0 \;=\; \bar{c} \;=\; \text{constant}\,, \tag{9.21}$$

with c being the total concentration and $\bar{c}$ the corresponding equilibrium value, $\lim_{t\to\infty} c(t) = \bar{c}$.


0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
$P_n(t)$
0
50
100
150
200
n
0.030
0.025
0.020
0.015
0.010
0.005
0.000
$P_n(t)$
0
500
1000
1500
2000
n
0.010
0.008
0.006
0.004
0.002
0.000
$P_n(t)$
0
5000
10 000
15 000
20 000
n

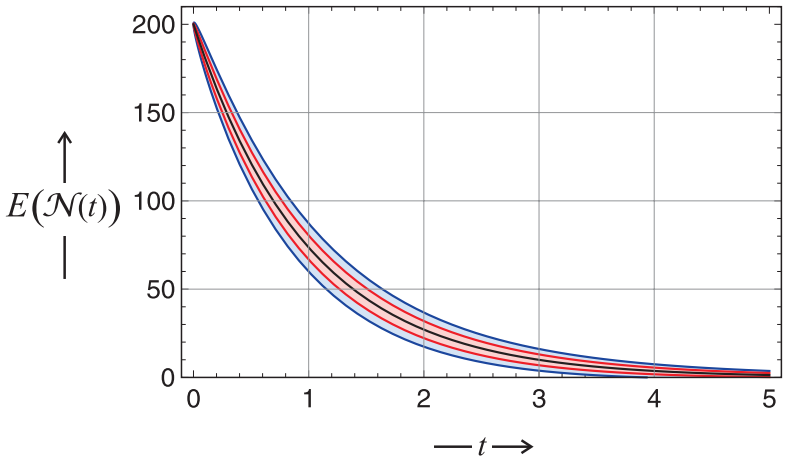


Figure 9.4: **Probability density of an irreversible monomolecular reaction.** The three plots on the previous page show the evolution of the probability density, $P_n(t)$, of the number of molecules of a compound **A** which undergo a reaction **A**→**B**. The initially infinitely sharp density $P_n(0) = \delta_{n,n_0}$ becomes broader with time until the variance reaches its maximum at time $t = t_{1/2} = \ln 2/k$ 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 $E\big(\mathcal{N}_A(t)\big)$ and the confidence intervals $E \pm \sigma$ (68,3%,red) and $\pm 2\sigma$ (95,4%,blue) with $\sigma^2\big(\mathcal{N}_A(t)\big)$ 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 irreversible monomolecular reaction. We start by discussing the simpler irreversible case,

$$\mathbf{A} \xrightarrow{\;k\;} \mathbf{B}\,, \tag{9.19a'}$$

which is can be modeled and analyzed in full analogy to the previous case of the flow equilibrium. Although we are dealing with two molecular species, **A** and **B** the process is described by a single stochastic variable, $\mathcal{N}_A(t)$, since we have $\mathcal{N}_B(t) = n_0 - \mathcal{N}_A(t)$ with $n_0 = n(0)$ being the number of **A** molecules initially present because of the conservation relation (9.21). If a sufficiently small time interval is applied, the irreversible monomolecular reaction is modeled by a single step birth-and-death process with $w^+(n) = 0$

and $w^-(n) = kn$.[6] The probability density is defined by $P_n(t) = P(\mathcal{N}_A = n)$ and its time dependence obeys

$$\frac{\partial P_n(t)}{\partial t} \;=\; k\,(n+1)\,P_{n+1}(t) \;-\; k\,n\,P_n(t)\;. \tag{9.22}$$

The master equation (9.22) is solved again by means of the probability generating function,

$$g(s,t) \;=\; \sum_{n=0}^{\infty} P_n(t)\,s^n\,;\;\; |s| \leq 1\;,$$

which is determined by the PDE

$$\frac{\partial g(s,t)}{\partial t} \;-\; k\,(1-s)\,\frac{\partial g(s,t)}{\partial s} \;=\; 0.$$

The computation of the characteristic manifold of this PDE is tantamount to solving the ODE

$$k\;\mathrm{d}t \;=\; \frac{\mathrm{d}s}{s-1} \quad\Longrightarrow\quad e^{kt} \;=\; s - 1 \;+\; \text{const}\;.$$

With $\phi(s,t) = (s-1)\exp(-kt) + \gamma$, $g(s,t) = f(\phi)$, the normalization condition $g(1,t) = 1$, and the boundary condition $g(s,0) = f(\phi)_{t=0} = s^{n_0}$ we find

$$g(s,t) \;=\; \Bigl(s\cdot e^{-kt} \;+\; 1 \;-\; e^{-kt}\Bigr)^{n_0}\;. \tag{9.23}$$

This expression is easily expanded in binomial form, which orders with respect to increasing powers of s,

$$\begin{aligned} g(s,t) = (1-e^{-kt})^{n_0} + \binom{n_0}{1} s e^{-kt}(1-e^{-kt})^{n_0-1} + \binom{n_0}{2} s e^{-2kt}(1-e^{-kt})^{n_0-2} + \\ + \ldots + \binom{n_0}{n_0-1} s^{n_0-1} e^{-(n_0-1)kt}(1-e^{-kt}) + s^{n_0} e^{-n_0 kt}\;. \end{aligned}$$

Comparison of coefficients yields the time dependent probability density

$$P_n(t) \;=\; \binom{n_0}{n} \Bigl(\exp(-kt)\Bigr)^n \Bigl(1-\exp(-kt)\Bigr)^{n_0-n}\;. \tag{9.24}$$

[6] We remark that $w^-(0) = 0$ and $w^+(0) = 0$ are fulfilled, which are the conditions for a natural *absorbing barrier* at $n = 0$.

It is straightforward to compute the expectation value of the stochastic variable $\mathcal{N}_A$, which coincides again with the deterministic solution, and its variance

$$\begin{aligned} E\big(\mathcal{N}_A(t)\big) &= n_0\, e^{-kt}\ , \\ \sigma^2\big(\mathcal{N}_A(t)\big) &= n_0\, e^{-kt}\,\big(1-e^{-kt}\big)\ . \end{aligned} \tag{9.25}$$

The half-life of a population of n_0 particles,

$$t_{1/2}\ :\ E\{\mathcal{N}_A(t)\} = \frac{n_0}{2} = n_0\cdot e^{-kt_m} \quad\Longrightarrow\quad t_{1/2} = \frac{1}{k}\ln 2\ ,$$

is time of maximum variance or standard deviation, $\mathrm{d}\sigma^2/\,\mathrm{dt} = 0$ or $\mathrm{d}\sigma/\,\mathrm{dt} = 0$, respectively. An example of the time course of the probability density of an irreversible monomolecular reaction is shown in Fig. 9.4.

The reversible monomolecular reaction. The analysis of the irreversible reaction is readily extended to the reversible case (9.19), where we are again dealing with a one step birth-and-death process in a closed system: The conservation relation $\mathcal{N}_A(t)+\mathcal{N}_B(t)=n_0$ – with n_0 being again the number of molecules of class **A** initially present, $P_n(0)=\delta_{n,n0}$ – holds and the transition probabilities are given by: $w^+(n)=k_2(n_0-n)$ and $w^-(n)=k_1 n$.[7] The master equation is now of the form

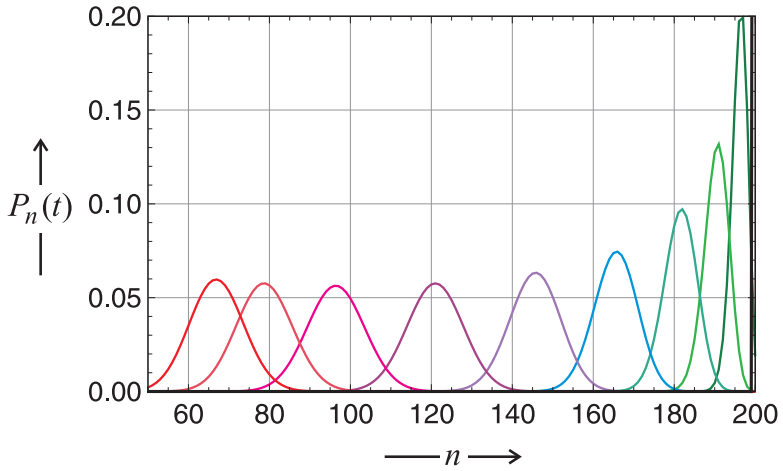
$$\begin{aligned} \frac{\partial P_n(t)}{\partial t} &= k_2(n_0-n+1)P_{n-1}(t) + k_1(n+1)P_{n+1}(t) - \\ &\quad - \Big(k_1 n + k_2(n_0-n)\Big)P_n(t)\ . \end{aligned} \tag{9.26}$$

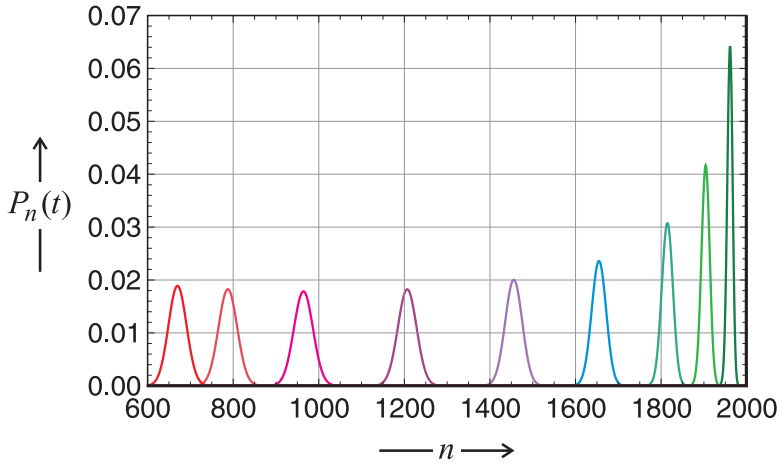
Making use of the probability generating function $g(s,t)$ we derive the PDE

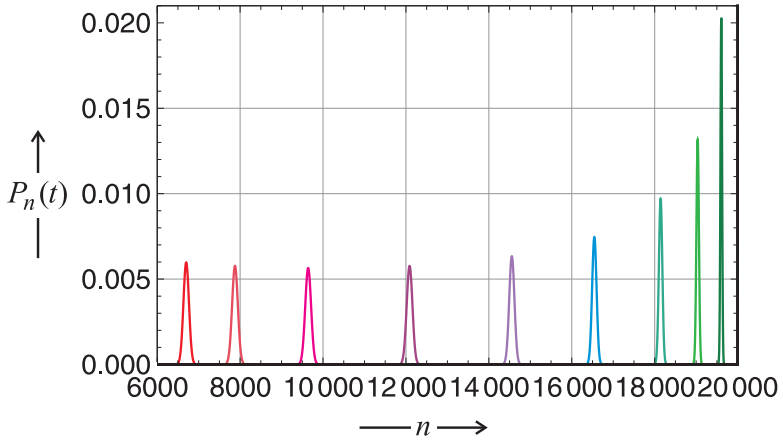
$$\frac{\partial g(s,t)}{\partial t} = \Big(k_1+(k_2-k_1)s-k_1 s^2\Big)\frac{\partial g(s,t)}{\partial s} + n_0\,k_2(s-1)\,g(s,t)\ .$$

The solutions of the PDE are simpler when expressed in terms of parameter combinations, $\kappa=k_1+k_2$ and $\lambda=k_1/k_2$, and the function

[7] Here we note the existence of barriers at $n=0$ and $n=n_0$, which are characterized by $w^-(0)=0$, $w^+(0)=k_2 n_0>0$ and $w^+(n_0)=0$, $w^-(n_0)=k_1 n_0>0$, respectively. These equations fulfil the conditions for *reflecting barriers*.


0.20
0.15
$P_n(t)$
0.10
0.05
0.00
60
80
100
120
140
160
180
200
n


0.07
0.06
0.05
$P_n(t)$
0.04
0.03
0.02
0.01
0.00
600
800
1000
1200
1400
1600
1800
2000
n


0.020
0.015
$P_n(t)$
0.010
0.005
0.000
6000
8000
10 000
12 000
14 000
16 000
18 000
20 000
n

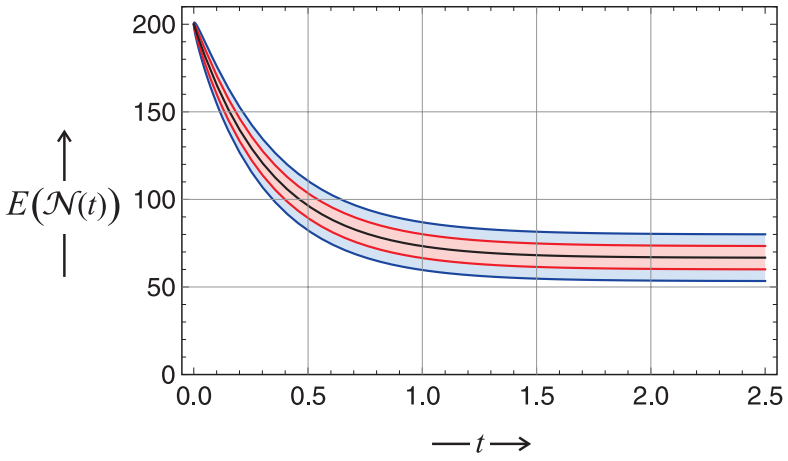


Figure 9.5: **Probability density of a reversible monomolecular reaction.** The three plots on the previous page show the evolution of the probability density, $P_n(t)$, of the number of molecules of a compound **A** which undergo a reaction **A**$\rightleftharpoons$**B**. The initially infinitely sharp 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 $E\big(\mathcal{N}_A(t)\big)$ and the confidence intervals $E\pm\sigma$ (68,3%,red) and $\pm 2\sigma$ (95,4%,blue) with $\sigma^2\big(\mathcal{N}_A(t)\big)$ being the variance. Parameters used: $n_0 = 200$, 2000, and 20 000; $k_1 = 2\,k_2 = 1\,[t^{-1}]$; 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).

$\omega(t) = \lambda \exp(-\kappa t) + 1$:

$$\begin{aligned} g(s,t) &= \left(1 + (s-1)\,e^{-\kappa t} - \frac{s}{\lambda}\right)^{n_0} = \\ &= \left(\frac{\lambda\,(1-e^{-\kappa t}) + s\,(\lambda e^{-\kappa t}+1)}{1+\lambda}\right)^{n_0} = \\ &= \sum_{n=0}^{n_0} \left(\binom{n_0}{n} \left(\lambda e^{-\kappa t}+1\right)^n \left(\lambda(1-e^{-\kappa t})\right)\right)^{n_0-n} \frac{s^n}{(1+\lambda)^{n_0}}\,. \end{aligned}$$

The probability density for the reversible reaction is then obtained as

$$P_n(t) = \binom{n_0}{n} \frac{1}{(1+\lambda)^{n_0}} \left(\lambda e^{-\kappa t}+1\right)^n \left(\lambda(1-e^{-\kappa t})\right)^{n_0-n}\,. \qquad (9.27)$$

Expectation value and variance of the numbers of molecules are readily com-

puted (with $\omega(t) = \lambda \exp(-\kappa t) + 1$):

$$\begin{aligned} E\big(\mathcal{N}_A(t)\big) &= \frac{n_0}{1+\lambda}\,\omega(t)\ , \\ \sigma^2\big(\mathcal{N}_A(t)\big) &= \frac{n_0\,\omega(t)}{1+\lambda}\left(1 - \frac{\omega(t)}{1+\lambda}\right)\ , \end{aligned} \tag{9.28}$$

and the stationary values are

$$\begin{aligned} \lim_{t\to\infty} E\big(\mathcal{N}_A(t)\big) &= n_0\,\frac{k_2}{k_1+k_2}\ , \\ \lim_{t\to\infty} \sigma^2\big(\mathcal{N}_A(t)\big) &= n_0\,\frac{k_1\,k_2}{(k_1+k_2)^2}\ , \\ \lim_{t\to\infty} \sigma\big(\mathcal{N}_A(t)\big) &= \sqrt{n_0}\,\frac{\sqrt{k_1\,k_2}}{k_1+k_2}\ . \end{aligned} \tag{9.29}$$

This result shows that the $\sqrt{N}$-law is fulfilled up to a factor that is independent of N: $E/\sigma = \sqrt{n_0}\,k_2/\sqrt{k_1\,k_2}$.

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, $\bar{\sigma^2} = k_1 k_2\, n_0/(k_1+k_2)^2$. The time of maximal fluctuations, $t_{\max}$, is easily calculated from the condition $\mathrm{d}\sigma^2/\,\mathrm{dt} = 0$ and one obtains

$$t_{\max} = \frac{1}{k_1+k_2}\,\ln\left(\frac{2\,k_1}{k_1-k_2}\right)\ . \tag{9.30}$$

Depending on the sign of $(k_1 - k_2)$ 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 Fig. 9.5 where a case with $k_1 > k_2$ is presented.

In order to illustrate fluctuations and their value under equilibrium conditions the Austrian physicist Paul Ehrenfest designed a game called Ehrenfest's urn model [64], 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.

Two examples of bimolecular reactions (9.1f) and (9.1j) as well as their solutions are presented here in order to illustrate the enormous degree of sophistication that is required to derive analytical solutions,

$$\mathbf{A} + \mathbf{B} \xrightarrow{k} \mathbf{C} \quad \text{and} \tag{9.31a}$$

$$2\,\mathbf{A} \xrightarrow{k} \mathbf{B}\,, \tag{9.31b}$$

and we discuss them in this sequence.

The irreversible bimolecular addition reaction. As first example of a bimolecular process we choose the simple irreversible bimolecular addition reaction (9.31a). In this case we are dealing with three dependent stochastic variables $\mathcal{N}_A(t)$, $\mathcal{N}_B(t)$, and $\mathcal{N}_C(t)$. Following McQuarrie we define the probability $P_n(t) = P\big(\mathcal{N}_A(t) = n\big)$ and apply the standard initial condition $P_n(0) = \delta_{n,n_0}$, $P(\mathcal{N}_B(0) = b) = \delta_{b,b_0}$, and $P(\mathcal{N}_C(0) = c) = \delta_{c,0}$. Accordingly, we have from the laws of stoichiometry $\mathcal{N}_B(t) = b_0 - n_0 + \mathcal{N}_A(t)$ and $\mathcal{N}_C(t) = n_0 - \mathcal{N}_A(t)$. For simplicity we denote $b_0 - n_0 = \Delta_0$. Then the master equation for the chemical reaction is of the form

$$\frac{\partial P_n(t)}{\partial t} \;=\; k\,(n+1)\,(\Delta_0 + n + 1)\,P_{n+1}(t) \;-\; k\,n\,(\Delta_0 + n)\,P_n(t)\;. \tag{9.32}$$

We remark that the birth and death rates are no longer linear in n. The corresponding PDE for the generating function is readily calculated

$$\frac{\partial g(s,t)}{\partial t} \;=\; k\,(\Delta_0 + 1)(1-s)\frac{\partial g(s,t)}{\partial s} + k\,s(1-s)\,\frac{\partial^2 g(s,t)}{\partial s^2}\;. \tag{9.33}$$

The derivation of solutions or this PDE is quite demanding. It can be achieved by separation of variables:

$$g(s,t) \;=\; \sum_{m=0}^{\infty} A_m\,Z_m(s)\,T_m(t)\;. \tag{9.34}$$

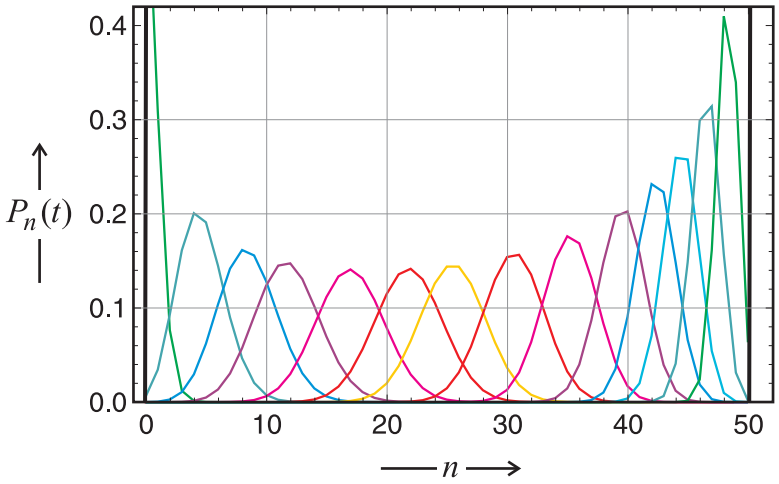


Figure 9.6: **Irreversible bimolecular addition reaction A+B → C.** The plot shows the probability distribution $P_n(t) = \text{Prob}\big(\mathcal{N}_C(t) = n\big)$ describing the number of molecules of species **C** as a function of time and calculated by equation (9.39). The initial conditions are chosen to be $\mathcal{N}_A(t) = \delta(a, a_0)$, $\mathcal{N}_B(t) = \delta(b, b_0)$, and $\mathcal{N}_C(t) = \delta(c, 0)$. 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: $\lim_{t\to\infty} \mathcal{N}_C(t) = \delta\big(n, \min(a_0, b_0)\big)$. 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).

We dispense from details and list only the coefficients and functions of the solution:

$$
\begin{aligned}
A_m &= (-1)^m \, \frac{(2m + \Delta_0)\Gamma(m + \Delta_0)\Gamma(n_0 + 1)\Gamma(n_0 + \Delta_0 + 1)}{\Gamma(m + 1)\Gamma(\Delta_0 + 1)\Gamma(n_0 - m + 1)\Gamma(n_0 + \Delta_0 + m + 1)} \;, \\
Z_m(s) &= J_m(\Delta_0, \Delta_0 + 1, s) \;, \quad \text{and} \\
T_m(t) &= \exp\big(-m(m + \Delta_0)\, kt\big) \;.
\end{aligned}
$$

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

$$s(1-s)\frac{\mathrm{d}^2 J_n(p,q,s)}{\mathrm{ds}^2} + \Big(q-(p+1)s\Big)\frac{\mathrm{d}J_n(p,q,s)}{\mathrm{ds}} + n(n+p)\,J_n(p,q,s) \;=\; 0\,.$$

These polynomials fulfil the following conditions:

$$\frac{\mathrm{d}J_n(p,q,s)}{\mathrm{ds}} \;=\; -\,\frac{n(n+p)}{s}\,J_{n-1}(p+2,q+1,s) \quad \text{and}$$

$$\int_0^1 s^{q-1}(1-s)^{p-q}J_n(p,q,s)\,J_\ell(p,q,s)\,ds \;=\; \frac{n!\Big(\Gamma(q)\Big)^2\Gamma(n+p-q+1)}{(2n+p)\Gamma(n+p)\Gamma(n+q)}\,\delta_{\ell,n}\,.$$

At the relevant value of the auxiliary variable, $s=1$, we differentiate twice and find:

$$\left(\frac{\partial g(s,t)}{\partial s}\right)_{s=1} = \sum_{m=1}^{n_0}\frac{(2m+\Delta_0)\Gamma(n_0+1)\Gamma(n_0+\Delta_0+1)}{\Gamma(n_0-m+1)\Gamma(n_0+\Delta_0+m+1)}\,T_m(t)\,, \tag{9.35}$$

$$\left(\frac{\partial^2 g(s,t)}{\partial s^2}\right)_{s=1} =$$

$$= \sum_{m=2}^{n_0}\frac{(m-1)(m+\Delta_0+1)(2m+\Delta_0)\Gamma(n_0+1)\Gamma(n_0+\Delta_0+1)}{\Gamma(n_0-m+1)\Gamma(n_0-\Delta_0+m+1)}\,T_m(t) \tag{9.36}$$

from which we obtain expectation value and variance

$$E\big(\mathcal{N}_A(t)\big) \;=\; \left(\frac{\partial g(s,t)}{\partial s}\right)_{s=1} \quad \text{and}$$

$$\sigma^2\big(\mathcal{N}_A(t)\big) \;=\; \left(\frac{\partial^2 g(s,t)}{\partial s^2}\right)_{s=1} + \left(\frac{\partial g(s,t)}{\partial s}\right)_{s=1} - \left(\left(\frac{\partial g(s,t)}{\partial s}\right)_{s=1}\right)^2. \tag{9.37}$$

As we see in the current example, bimolecularity complicates the solution of the chemical master equations substantially and makes analytical 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, $|\Delta_0| \gg n_0 > 1$, which is known as *pseudo first order condition*. Then, the sums can be approximated well be the first terms and we find (with $k' = \Delta_0 k$):

$$\left(\frac{\partial g(s,t)}{\partial s}\right)_{s=1} \;\approx\; n_0\,\frac{\Delta_0+2}{n_0+\Delta_0+1}\,e^{-(\Delta_0+1)kt} \;\approx\; n_0\,e^{-k't} \quad \text{and}$$

$$\left(\frac{\partial^2 g(s,t)}{\partial s^2}\right)_{s=1} \;\approx\; n_0\,(n_0-1)\,e^{-2\,k't}\,,$$

and we obtain finally,

$$\begin{aligned} E\big(\mathcal{N}_A(t)\big) &= n_0\, e^{-k't} \quad \text{and} \\ \sigma^2\big(\mathcal{N}_A(t)\big) &= n_0\, e^{-k't}\, \Big(1 - e^{-k't}\Big)\ , \end{aligned} \tag{9.38}$$

which is essentially the same result as obtained for the irreversible first order reaction.

For the calculation of the probability density we make use of a slightly different definition of the stochastic variables and use $\mathcal{N}_C(t)$ counting the number of molecules **C** in the system: $P_n(t) = P\big(\mathcal{N}_C(t) = n\big)$. With the initial condition $P_n(0) = \delta(n,0)$ and the upper limit of n, $\lim_{t\to\infty} P_n(t) = c$ with $c = \min\{a_0, b_0\}$ where a_0 and b_0 are the sharply defined numbers of **A** and **B** molecules initially present ($\mathcal{N}_A(0) = a_0$, $\mathcal{N}_B(0) = b_0$), we have

$$\sum_{n=0}^{c} P_n(t) = 1 \ \text{ and thus } \ P_n(t) = 0 \ \forall\ (n \notin [0,c], n \in \mathbb{Z})$$

and the master equation is now of the form

$$\begin{aligned} \frac{\partial P_n(t)}{\partial t} &= k\,\big(a_0 - (n-1)\big)\big(b_0 - (n-1)\big)P_{n-1}(t) - \\ &\quad - k\,(a_0 - n)(b_0 - n)P_n(t)\ . \end{aligned} \tag{9.32'}$$

In order to solve the master equation (9.32') the probability distribution $P_n(t)$ is Laplace transformed in order to obtain a set of pure difference equation from the master equation being a set of differential-difference equation

$$q_n(s) \ = \ \int_0^\infty \exp(-\,s\cdot t)\, P_n(t)\,\mathrm{dt}$$

and with the initial condition $P_n(0) = \delta(n,0)$ we obtain

$$\begin{aligned} -1 \,+\, s\, q_0(s) &= -\,k\, a_0\, b_0\, q_0(s)\ , \\ s\, q_n(s) &= k\,\big(a_0 - (n-1)\big)\big(b_0 - (n-1)\big)\, q_{n-1}(s) - \\ &\quad - k\,(a_0 - n)(b_0 - n)\, q_n(s)\,,\ \ 1 \le n \le c\ . \end{aligned}$$

Successive iteration yields the solutions in terms of the functions $q_n(s)$

$$q_n(s) \ = \ \binom{a_0}{n}\binom{b_0}{n}(n!)^2 k^n \prod_{j=0}^{n} \frac{1}{s + k(a_0 - j)(b_0 - j)}\,,\ 0 \le n \le c$$

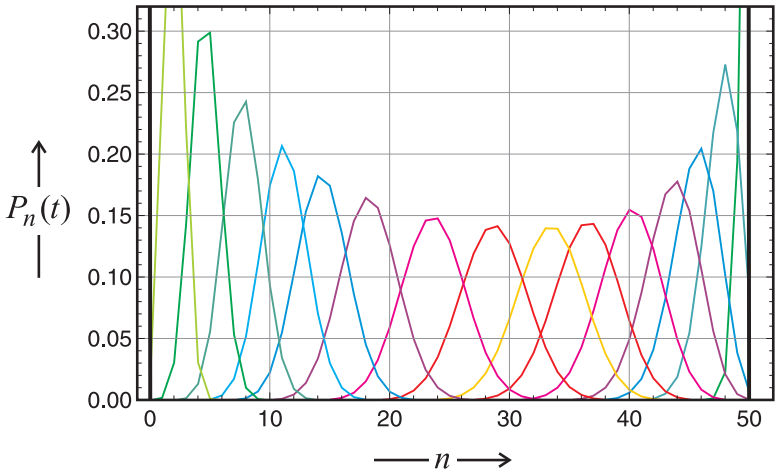


Figure 9.7: **Irreversible dimerization reaction $2\mathbf{A} \rightarrow \mathbf{C}$.** The plot shows the probability distribution $P_n(t) = \text{Prob}\big(\mathcal{N}_A(t) = n\big)$ describing the number of molecules of species **C** as a function of time and calculated by equation (9.45). The number of molecules **C** is given by the distribution $P_m(t) = \text{Prob}\big(\mathcal{N}_C(t) = m\big)$. The initial conditions are chosen to be $\mathcal{N}_A(t) = \delta(n, a_0)$, and $\mathcal{N}_C(t) = \delta(m, 0)$ 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: $\lim_{t\to\infty} \mathcal{N}_A(t) = \delta(n, 0)$ $\lim_{t\to\infty} \mathcal{N}_C(t) = \delta(m, a_0/2)$. 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).

and after converting the product into partial fractions and inverse transformation one finds the result

$$\begin{aligned} P_n(t) \;=\; (-1)^n \binom{a_0}{n} \binom{b_0}{n} \sum_{j=0}^{n} (-1)^j \left(1 + \frac{n-j}{a_0 + b_0 - n - j}\right) \times \\ \times \binom{n}{j} \binom{a_0 + b_0 - j}{n}^{-1} e^{-k(a_0-j)(b_0-j)t} \,. \end{aligned} \tag{9.39}$$

An illustrative example is shown in Fig. 9.6. The difference between the irreversible reactions monomolecular conversion and the bimolecular addition reaction (Fig. 9.4) is indeed not spectacular.

The dimerization reaction. When the dimerization reaction (9.1j) is modeled by means of a master equation [203] we have to take into account that two molecules **A** vanish at a time, and an individual jump involves always $\Delta n = 2$:

$$\frac{\partial P_n(t)}{\partial t} \;=\; \frac{1}{2}\,k\,(n+2)(n+1)\,P_{n+2}(t) \;-\; \frac{1}{2}\,k\,n(n-1)\,P_n(t)\;, \tag{9.40}$$

which gives rise to the following PDE for the probability generating function

$$\frac{\partial g(s,t)}{\partial t} \;=\; \frac{k}{2}\,(1-s^2)\,\frac{\partial^2 g(s,t)}{\partial s^2}\;. \tag{9.41}$$

The analysis of this PDE is more involved than it might look at a first glance. Nevertheless, an exact solution similar to (9.34) is available:

$$g(s,t) \;=\; \sum_{m=0}^{\infty} A_m\,C_m^{-\frac{1}{2}}(s)\,T_m(t)\;, \tag{9.42}$$

wherein the parameters and functions are defined by

$$\begin{aligned}
A_m \;&=\; \frac{1-2m}{2^m}\cdot\frac{\Gamma(n_0+1)\,\Gamma[(n_0-m+1)/2]}{\Gamma(n_0-m+1)\,\Gamma[(n_0+m+1)/2]}\;,\\
C_m^{-\frac{1}{2}}(s)\;&:\;\;(1-s^2)\,\frac{\mathrm{d}^2 C_m^{-\frac{1}{2}}(s)}{\mathrm{ds}^2} \;+\; m(m-1)\,C_m^{-\frac{1}{2}}(s) \;=\; 0\;,\\
T_m(t) \;&=\; \exp\{-\frac{1}{2}\,k\,m(m-1)\,t\}\;.
\end{aligned}$$

The functions $C_m^{-\frac{1}{2}}(s)$ 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 $\mathcal{N}_A(t)$ (μ stands for an integer running index, $\mu \in \mathbb{N}$):

$$\begin{aligned}
E\big(\mathcal{N}_A(t)\big) \;&=\; -\sum_{m=2\mu=2}^{2\lfloor\frac{n_0}{2}\rfloor} A_m\,T_m(t) \quad \text{and}\\
\sigma^2\big(\mathcal{N}_A(t)\big) \;&=\; -\sum_{m=2\mu=2}^{2\lfloor\frac{n_0}{2}\rfloor} \Big(\frac{1}{2}(m^2-m+2)\,A_m\,T_m(t) \;+\; A_m^2\,T_m^2(t)\Big)\;.
\end{aligned} \tag{9.43}$$

In order to obtain concrete results these expressions can be readily evaluated numerically.

There is one interesting detail in the deterministic version of the dimerization reaction. It is conventionally modeled by the differential equation (9.44a), which can be solved readily. The correct ansatz, however, would be (9.44b) for which we have also an exact solution (with $[\mathbf{A}]=a(t)$ and $a(0)=a_0$):

$$-\frac{\mathrm{d}a}{\mathrm{dt}} = k\,a^2 \quad\Longrightarrow\quad a(t) = \frac{a_0}{1+a_0\,kt} \quad \text{and} \tag{9.44a}$$

$$-\frac{\mathrm{d}a}{\mathrm{dt}} = k\,a(a-1) \quad\Longrightarrow\quad a(t) = \frac{a_0}{a_0+(1-a_0)e^{-kt}}\;. \tag{9.44b}$$

The expectation value of the stochastic solution lies always between the solution curves (9.44a) and (9.44b). An illustrative example is shown in figure 9.7.

As the previous paragraph we consider also a solution of the master equation by means of a Laplace transformation [158]. Since we are dealing with a step size of two molecules **A** converted into one molecule **C**, the master equation is defined only for odd or only for even numbers of molecules **A**. For an initial number of $2a_0$ molecules and a probability $P_2n(t) = P\big(\mathcal{N}_A(t)=2n\big)$ we have for the initial conditions $\mathcal{N}_A(0)=2a_0$, $\mathcal{N}_C(0)=0$ and the condition that all probabilities outside the interval $[0,2a_0]$ as well as the odd probabilities P_{2n-1} $(n=1,\ldots,2a_0-1)$ vanish

$$\frac{\partial P_{2n}(t)}{\partial t} = -\frac{1}{2}\,k\,(2n)(2n-1)P_{2n}(t) + \frac{1}{2}\,k\,(2n+2)(2n+1)P_{2n+2}(t) \tag{9.40'}$$

The probability distribution $P_{2n}(t)$ is derived as in the previous subsection by Laplace transformation

$$q_{2y}(s) = \int_0^\infty \exp(-\,s\cdot t)\,P_{2y}(t)\;\mathrm{dt}$$

yielding the set of difference equations

$$\begin{aligned}
-1 + s\,q_{2a_0}(s) &= -\frac{1}{2}\,k\,(2a_0)(2a_0-1)\,q_{2a_0}(s)\,,\\
s\,q_{2n}(s) &= -\frac{1}{2}\,k\,(2n)(2n-1)\,q_{2n}(s) + \\
&\quad + \frac{1}{2}\,k\,(2n+2)(2n+1)\,q_{2n+2}(s)\,,\;\; 0\le y\le a_0-1\;,
\end{aligned}$$

which again can be solved by successive iteration. It is straightforward to calculate first the Laplace transform for 2μ, the number of molecules of species **A** that have reacted to yield **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:

$$\begin{aligned} P_{2(a_0-m)}(t) \;&=\; (-1)^m \frac{a_0!\,(2a_0-1)!!}{(a-m)!\,(2a_0-2m-1)} \times \\ &\times \sum_{j=0}^{m} (-1)^j \frac{(4a_0-4j-1)(4a_0-2m-2j-3)!!}{j!(m-j)!(4a_0-2j-1)!!} \times \\ &\times\; e^{-k\,(a_0-j)\cdot\left(2(a_0-j)-1\right)\,t} . \end{aligned}$$

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

$$\begin{aligned} P_{2n}(t) \;&=\; (-1)^n \frac{a_0!(2a_0-1)!!}{n!(2n-1)!!} \times \\ &\times \sum_{i=1}^{n} (-1)^i \frac{(4i-1)(2n+2i-3)!!}{n!(2n-1)!!} \;\times\; e^{-k\,i(2i-1)t} . \end{aligned} \tag{9.45}$$

The results are illustrated be means of a numerical example in figure 9.7.

The examples discussed here provide detailed information on the analytically analyzable cases. The majority of systems, however, is to complicated for the analytical approach and here numerical simulation has to fill the gap

between model building and exact analysis. One major result concerns the double role of nonlinearity in chemistry. Like in deterministic reaction kinetics we encounter substantial complication in the mathematical handling of bi- and higher molecular processes, which do not manifest themselves in spectacular qualitative changes in reaction dynamics. On the other hand nonlinearities based on autocatalytic reaction steps may have rather drastic influence on the appearance of the reactions.

9.3 Computer simulation of master equations

In this section we introduce a model for computer simulation of stochastic chemical kinetics that has been developed and put upon a solid basis by the American physicist and mathematical chemist Daniel Gillespie [112, 113, 115, 117]. Considered is a population of N molecular species, $\{\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_N\}$ in the gas phase, which interact through M elementary chemical reactions $(\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M)$.[8] Two conditions are assumed to be fulfilled by the system: (i) the container with constant volume V in the sense of a flow reactor (CSTR in Fig. 4.3) is assumed to be *well mixed* by efficient stirring and (ii) the system is assumed to be in *thermal equilibrium* at constant temperature T. The goals of the simulation are the computation of the time course of the stochastic variables – $\mathcal{X}_k(t)$ being the number of molecules ($\mathbf{K}$) of species $\mathbf{S}_k$ at time t – and the description of the evolution of the population. A single computation yields a single trajectory, very much in the sense of a single solution of a stochastic differential equation (Fig. 8.18) and observable results are commonly derived through sampling of a sufficiently large number of trajectories.

For a reaction mechanism involving N species in M reactions the entire population is characterized by an N-dimensional random vector counting numbers of molecules for the various species $\mathbf{S}_k$,

$$\vec{\mathcal{X}}(t) \;=\; \big(\mathcal{X}_1(t), \mathcal{X}_2(t), \ldots, \mathcal{X}_N(t)\big)\;. \tag{9.46}$$

The common variables in chemistry are concentrations rather than particle

[8] Elementary steps of chemical reactions were defined and discussed in subsection 9.1.1.

numbers:

$$\mathbf{x} \;=\; \big(\mathrm{x}_1(t), \mathrm{x}_2(t), \ldots, \mathrm{x}_N(t)\big) \;\text{ with }\; \mathrm{x}_k \;=\; \frac{\mathcal{X}_k}{V \cdot N_L}, \tag{9.47}$$

where the volume V is the appropriate expansion parameter Ω for the system size.[9] The following derivation of the chemical master equation [115, pp. 407-417] focuses on reaction channels $\mathbf{R}_{\boldsymbol{\mu}}$ of bimolecular nature

$$\mathsf{S}_a \;+\; \mathsf{S}_b \;\longrightarrow\; \mathsf{S}_c \;+\; \ldots \tag{9.48}$$

like (9.1f,9.1i,9.1j and 9.1k) shown in the list (9.1). An extension to monomolecular and trimolecular reaction channels is straightforward, and *zero-molecular* processes like the influx of material into the reactor in the elementary step (9.1a) provide no major problems. Reversible reactions, for example (9.19), are handled as two elementary steps, $\mathsf{A} + \mathsf{B} \longrightarrow \mathsf{C} + \mathsf{D}$ and $\mathsf{C} + \mathsf{D} \longrightarrow \mathsf{A} + \mathsf{B}$. In equation (9.48) we distinguish between *reactant* species, A and B, and *product* species, C ... , of a reaction $\mathbf{R}_{\boldsymbol{\mu}}$.

9.3.1 Master equations from collision theory

Molecules in the vapor phase change their directions of flight irregularly through collisions. Collisions in the gaseous state are classified as elastic, inelastic, and reactive collisions. Because of the large numbers of molecules in the reaction volume the collisions occur – for all practical implications – independently of previous collisions, and hence follow a Poisson distribution (subsections 8.1.3 and 9.1.2) with the mean number of reactive collisions per time interval or reaction rate r as parameter $\lambda = r$. The reaction rate for reactive collisions of molecules A and B can be obtained from statistical mechanics of collisions in an ideal gas mixture and is given by[10]

$$\frac{\mathrm{d}[\mathsf{A}]}{\mathrm{dt}} \;=\; \frac{\mathrm{d}[\mathsf{B}]}{\mathrm{dt}} \;=\; r \;=\; Z\,\rho\,[\mathsf{A}]\,[\mathsf{B}]\,\exp\left(-\frac{E_a}{RT}\right)\;,$$

[9] In order to distinguish random and deterministic variables, stochastic concentrations are indicated by *upright* fonts.

[10] The mass action relevant variable of compound A is denoted by $[\mathsf{A}]$. It may be a particle number, N_{A}, a partial pressure p_{A} typical for the vapor phase, or a concentration c_{A} or activity a_{A} in solution.

where ρ is a steric factor, E_a is the energy of activation of the reaction, T is the absolute temperature, R is the gas constant and Z is the frequency of collisions between molecules **A** and **B**,

$$Z \;=\; \sigma_{\mathbf{AB}} \sqrt{\frac{8k_{\mathrm{B}}\,T}{\pi\,\mu_{\mathbf{AB}}}}\,N_{\mathrm{L}} \;\;\text{with}\;\; \sigma_{\mathbf{AB}} \approx \pi(d_{\mathbf{A}}+d_{\mathbf{B}})^2 \;\;\text{and}\;\; \mu_{\mathbf{AB}} \;=\; \frac{m_{\mathbf{A}}\,m_{\mathbf{B}}}{m_{\mathbf{A}}+m_{\mathbf{B}}}\;.$$

Here, $\sigma_{\mathbf{AB}}$ is the reaction cross section that can be estimated from the diameter of the collision complex, which for hard spheres simply is the sum of the diameters of the molecules: $d_{\mathbf{AB}} = d_{\mathbf{A}} + d_{\mathbf{B}}$ and $\mu_{\mathbf{AB}}$ is the reduced mass (see Fig. 9.8). In solution collision are replaced by molecular encounters since molecules move by diffusion in a random walk like manner with very small steps. The basic equations remain the same but the individual parameters have to be interpreted completely differently from the vapor phase. Reactions in biochemistry follow essentially the same rate laws as in chemistry although special notations and approximations are used for enzyme catalyzed reactions.

The two stipulations (i) *perfect mixture* and (ii) *thermal equilibrium* can now be cast into precise physical meanings. Premise (i) requires that the probability of finding the 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) implies that the velocity of a randomly chosen molecule of mass m will be found to lie within an infinitesimal region $\mathrm{d}\boldsymbol{v}^3$ around the velocity $\boldsymbol{v}$ is equal to

$$P_{\mathrm{M}B} \;=\; \left(\frac{m}{2\pi k_{\mathrm{B}}T}\right)\,e^{-mv^2/(2k_{\mathrm{B}}T)}\;.$$

Here, the velocity vector is denoted by $\boldsymbol{v} = (v_x, v_y, v_z)$ in Cartesian coordinates, the infinitesimal volume element fulfils $\mathrm{d}\boldsymbol{v}^3 = \mathrm{d}v_x\,\mathrm{d}v_y\,\mathrm{d}v_z$, the square of the velocity is $v^2 = v_x^2 + v_y^2 + v_z^2$, and k_{B} is Boltzmann's constant. Premise (ii) asserts that the velocities of molecules follow a *Maxwell-Boltzmann distribution* or formally it states that each Cartesian velocity component of a

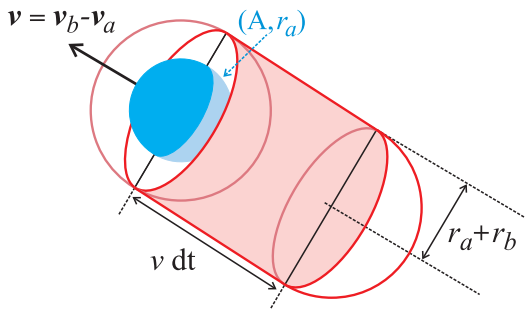


Figure 9.8: **Sketch of a molecular collision in dilute gases.** A spherical molecule $\mathbf{S}_a$ with radius r_a moves with a velocity $\boldsymbol{v} = \boldsymbol{v}_b - \boldsymbol{v}_a$ relative to a spherical molecule $\mathbf{S}_b$ with radius r_b. If the two molecules are to collide within the next infinitesimal time interval dt, the center of $\mathbf{S}_b$ has to lie inside a cylinder of radius $r = r_a + r_b$ and height $v\,\mathrm{dt}$. 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 the non-deformed one.

randomly selected molecule of mass m is represented by a random variable, which is normally distributed with mean 0 and variance $k_\mathrm{B}T/m$. Implicitly, the two stipulations assert that the molecular position and velocity components are all statistically independent of each other. For practical purposes, we expect premises (i) and (ii) to be valid for any dilute gas system at constant temperature in which *nonreactive* molecular collisions occur much more frequently than *reactive* molecular collisions.

In order to derive a chemical master equation for the population variables $\mathscr{X}_k(t)$ some properties of the probability $\pi_{\boldsymbol{\mu}}(t, \mathrm{dt})$ with $\boldsymbol{\mu} = 1, \ldots, M$ that a randomly selected combination of the reactant molecules for reaction $\mathbf{R}_{\boldsymbol{\mu}}$ at time t will react to yield products within the next infinitesimal time interval $[t, t+\mathrm{dt}[$. With the assumptions made in the previously virtually all chemical reaction channels fulfil the condition

$$\pi_{\boldsymbol{\mu}}(t, \mathrm{dt}) \;=\; \gamma_{\boldsymbol{\mu}}\,\mathrm{dt}\;, \tag{9.49}$$

where the *specific probability rate parameter* $\gamma_{\boldsymbol{\mu}}$ is independent of dt. First, we calculate the rate parameter for a general bimolecular reaction by means of classical collision theory and then extend briefly to mono- and trimolec-

ular reactions. Apart from the quantum mechanical approach the theory of collisions in dilute gases is the best developed microscopic model for chemical reactions and well suited for a rigorous derivation of the master equation from molecular motion and events.

Bimolecular reactions. The occurrence of a reaction $\mathbf{A}+\mathbf{B}$ has to be preceded by a collision of an $\mathbf{S}_a$ molecule with an $\mathbf{S}_b$ molecule, and first we shall calculate the probability of such a collision in the reaction volume V. For simplicity molecular species are regarded as spheres with specific masses and radii, for example m_a and r_a for $\mathbf{S}_a$, and m_b and r_b for $\mathbf{S}_b$, respectively. A collision occurs whenever the center-to center distance of the two molecules R_{AB} decreases to $(R_{AB})_{\min} = r_a + r_b$. Next we define the probability that a randomly selected pair of $\mathbf{R}_{\boldsymbol{\mu}}$ reactant molecules at time t will collide within the next infinitesimal time interval $[t, t+\,\mathrm{dt}[$ by $\pi^*_{\boldsymbol{\mu}}(t,\,\mathrm{dt})$ and calculate it from the Maxwell-Boltzmann distribution of molecular velocities according to Fig. 9.8.

The probability that a randomly selected pair of reactant molecules $\mathbf{R}_{\boldsymbol{\mu}}$, one molecule $\mathbf{S}_a$ and one molecule $\mathbf{S}_b$, has a relative velocity $\boldsymbol{v} = \boldsymbol{v}_b - \boldsymbol{v}_a$ lying in an infinitesimal volume element $\mathrm{d}\boldsymbol{v}^3$ about $\boldsymbol{v}$ at time t is denoted by $P(\boldsymbol{v}(t), \mathbf{R}_{\boldsymbol{\mu}})$ and can be readily obtained from kinetic theory of gases:

$$P(\boldsymbol{v}(t), \mathbf{R}_{\boldsymbol{\mu}}) \;=\; \left(\frac{\mu}{2\pi\, k_{\mathrm{B}} T}\right) \exp\bigl(-\mu v^2/(2k_{\mathrm{B}}T)\bigr)\, \mathrm{d}\boldsymbol{v}^3 \;.$$

Herein $v = |\boldsymbol{v}| = \sqrt{v_x^2 + v_y^2 + v_z^2}$ is the value of the relative velocity and $\mu = m_a m_b/(m_a + m_b)$ is the reduced mass of the two $\mathbf{R}_{\boldsymbol{\mu}}$ molecules. Two properties of the probabilities $P(\boldsymbol{v}(t), \mathbf{R}_{\boldsymbol{\mu}})$ for different velocities $\boldsymbol{v}$ are important:
(i) The elements in the set of all velocity combinations, $\{E_{\boldsymbol{v}(t),\mathbf{R}_{\boldsymbol{\mu}}}\}$ are mutually exclusive, and
(ii) they are collectively exhaustive since $\boldsymbol{v}$ is varied over the entire three dimensional velocity space.
Now we relate the probability $P(\boldsymbol{v}(t), \mathbf{R}_{\boldsymbol{\mu}})$ to a collision event E_{col} by calculating the conditional probability $P(E_{\mathrm{col}}(t,\,\mathrm{dt})|E_{\boldsymbol{v}(t),\mathbf{R}_{\boldsymbol{\mu}}})$. In Fig. 9.8 we sketch the geometry of the collision event between to randomly selected spherical

molecules S_a and S_b that is assumed to occur with an infinitesimal time interval dt:[11] A randomly selected molecule S_a moves along the vector $\boldsymbol{v}$ of the relative velocity $\boldsymbol{v}_b - \boldsymbol{v}_b$ between S_a and an also randomly selected molecule S_b. A collision between the molecules will take place in the interval $[t, t+$ dt if and only if the center of molecule S_b is inside the spherically distorted cylinder (Fig. 9.8) at time t. Thus $P(E_{\mathrm{col}}(t,\,\mathrm{dt})|E_{\boldsymbol{v}(t),\mathbf{R}_{\boldsymbol{\mu}}})$ is the probability that the center of a randomly selected S_b molecule moving with velocity $\boldsymbol{v}(t)$ relative to the randomly selected S_a molecule will be situated at time t with a certain subregion of V, which has a volume $V_{col} = v\,\mathrm{dt}\cdot\pi(r_a + r_b)^2$, and by scaling with the total volume V we obtain:[12]

$$P\big(E_{\mathrm{col}}(t,\,\mathrm{dt})|E_{\boldsymbol{v}(t),\mathbf{R}_{\boldsymbol{\mu}}}\big) \;=\; \frac{v(t)\,\mathrm{dt}\cdot\pi(r_a + r_b)^2}{V}\;. \tag{9.50}$$

By substitution and integration over the entire velocity space we can calculate the desired probability

$$\pi^*_{\boldsymbol{\mu}}(t,\,\mathrm{dt}) \;=\; \iiint\limits_{\boldsymbol{v}} \left(\frac{\mu}{2\pi\,k_{\mathrm{B}}T}\right) e^{-\mu v^2/(2k_{\mathrm{B}}T)}\cdot\frac{v(t)\,\mathrm{dt}\cdot\pi(r_a + r_b)^2}{V}\;\mathrm{d}\boldsymbol{v}^3\;.$$

Evaluation of the integral is straightforward and yields

$$\pi^*_{\boldsymbol{\mu}}(t,\,\mathrm{dt}) \;=\; \left(\frac{8\pi\,k_{\mathrm{B}}T}{V}\right)^{1/2}\;\frac{(r_a + r_b)^2}{\mu}\;\mathrm{dt}\;. \tag{9.51}$$

The first factor contains only constants and the macroscopic quantities, volume V and temperature T, whereas the molecular parameters, the radii r_a and r_b and the reduced mass μ.

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

[11] The absolute time t comes into play because the positions of the molecules, $\boldsymbol{r}_a$ and $\boldsymbol{r}_b$, and their velocities, $\boldsymbol{v}_a$ and $\boldsymbol{v}_b$, depend on t.

[12] Implicitly in the derivation we made use of the infinitesimally small size of dt. Only if the distance v dt is vanishingly small, the possibility of collisional interference of a third molecule can be neglected.

of independent probabilities we have

$$\pi_{\boldsymbol{\mu}}(t,\,\mathrm{dt}) \;=\; p_{\boldsymbol{\mu}}\,\pi^{*}_{\boldsymbol{\mu}}(t,\,\mathrm{dt})\;,$$

and with respect to equation (9.49) we find

$$\gamma_{\boldsymbol{\mu}} \;=\; p_{\boldsymbol{\mu}}\left(\frac{8\pi\,k_{\mathrm{B}}T}{V}\right)^{1/2}\;\frac{(r_a+r_b)^2}{\mu}\;. \tag{9.52}$$

As said before, it is crucial for the forthcoming analysis that $\gamma_{\boldsymbol{\mu}}$ is independent of dt and this will be the case if and only if $p_{\boldsymbol{\mu}}$ 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 [115, pp.413-417]. It has to be remarked, however, that the application of classical collision theory to molecular details of chemical reactions can be an illustrative and useful heuristic at best, because the molecular domain falls into the realm of quantum phenomena and any theory that aims at a derivation of reaction probabilities from first principles has to be built upon a quantum mechanical basis.

Monomolecular, trimolecular, and other reactions. A monomolecular reaction is of the form $\mathbf{A} \longrightarrow \mathbf{C}$ and describes the spontaneous conversion

$$\mathbf{S}_a \;\longrightarrow\; \mathbf{S}_c\;. \tag{9.53}$$

One molecule $\mathbf{S}_a$ is converted into one molecule $\mathbf{S}_c$. This reaction is different from a catalyzed conversion

$$\mathbf{S}_a \,+\, \mathbf{S}_b \;\longrightarrow\; \mathbf{S}_c \,+\, \mathbf{S}_b\;, \tag{9.48'}$$

where the conversion $\mathbf{A} \longrightarrow \mathbf{C}$ is initiated by a collision of an $\mathbf{A}$ molecule with a $\mathbf{B}$ molecule,[13] and a description as an ordinary bimolecular process is straightforward.

[13]The two reactions are related by rigorous thermodynamics: Whenever a catalyzed reaction is part of a mechanism an incorporation of the corresponding uncatalyzed process in the reaction mechanism is required in order to fulfill the requirements of thermodynamics.

The true monomolecular conversion (9.53) is driven by some quantum mechanical mechanism similar as in the case of radioactive decay of a nucleus. Time-dependent perturbation theory in quantum mechanics [209, 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 monomolecular conversion to be approximately of the form $\gamma_{\boldsymbol{\mu}}\,\mathrm{dt}$ with $\gamma_{\boldsymbol{\mu}}$ being independent of dt.

Trimolecular reactions of the form

$$\mathsf{S}_a + \mathsf{S}_b + \mathsf{S}_c \longrightarrow \mathsf{S}_d + \dots \tag{9.54}$$

should not be considered because collisions of three particles do not occur with a probability larger than of measure zero. There may be, however, special situations where approximations of complicated processes by trimolecular events is justified. One example is a set of three coupled reactions with four reactant molecules [114, pp.359-361] where is was shown that $\pi_{\boldsymbol{\mu}}(t, \mathrm{dt})$ is essentially linear in dt.

The last class of reaction to be considered here is no proper chemical reaction but an influx of material into the reactor. It is often denoted as a the zeroth order reaction (9.1a):

$$* \longrightarrow \mathsf{S}_a\ . \tag{9.55}$$

Here, the definition of the influx and the *efficient mixing* or homogeneity condition is helpful, because it guarantees that the number of molecules entering the homogeneous system is a constant and does not depend on dt.

9.3.2 Simulation of master equations

So far we succeeded to derive the fundamental fact that for each elementary reaction channel $\mathbf{R}_{\boldsymbol{\mu}}$ with $\boldsymbol{\mu} = 1, \dots, M$, which is accessible to the molecules

of a well-mixed and thermally equilibrated system in the gas phase or in solution, exists a scalar quantity $\gamma_{\boldsymbol{\mu}}$, which is independent of dt such that [115, p.418]

$$\begin{aligned}\gamma_{\boldsymbol{\mu}}\ \mathrm{dt}\ &= \text{probability that a randomly selected combination of}\\ &\quad \mathbf{R}_{\boldsymbol{\mu}} \text{ reactant molecules at time } t \text{ will react accordingly}\\ &\quad \text{in the next infinitesimal time interval } [t, t+\mathrm{dt}[\,.\end{aligned} \tag{9.56}$$

The specific probability rate constant, $\gamma_{\boldsymbol{\mu}}$ is one of three quantities that are required to fully characterize a particular reaction channel $\mathbf{R}_{\boldsymbol{\mu}}$. In addition we shall require a function $h_{\boldsymbol{\mu}}(\boldsymbol{n})$ where the vector $\boldsymbol{n} = (n_1, \ldots, n_n)'$ contains the exact numbers of all molecules at time t, $\vec{\mathcal{N}}(t) = \big(\mathcal{N}_1(t), \ldots, \mathcal{N}_N(t)\big)' = \boldsymbol{n}(t)$,

$$\begin{aligned}h_{\boldsymbol{\mu}}(\boldsymbol{n})\ &\equiv \text{the number of distinct combinations of } \mathbf{R}_{\boldsymbol{\mu}} \text{ reactant}\\ &\quad \text{molecules in the system when the numbers of molecules}\\ &\quad \mathbf{S}_k \text{ are exactly } n_k \text{ with } k = 1, \ldots, N\ ,\end{aligned} \tag{9.57}$$

and an $N \times M$ matrix of integers $\mathbf{S} = \{\nu_{k\boldsymbol{\mu}};\ k = 1, \ldots, N,\ \boldsymbol{\mu} = 1, \ldots, M\}$, where

$$\begin{aligned}\nu_{k\boldsymbol{\mu}}\ &\equiv \text{the change in the } \mathbf{S}_k \text{ molecular population caused by the}\\ &\quad \text{occurrence of one } \mathbf{R}_{\boldsymbol{\mu}} \text{ reaction.}\end{aligned} \tag{9.58}$$

The functions $h_{\boldsymbol{\mu}}(\boldsymbol{n})$ and the matrix $\mathbf{N}$ are readily deduced by inspection of the algebraic structure of the reaction channels. We illustrate by means of an example:

$$\begin{aligned}\mathbf{R}_1:\ \mathbf{S}_1 + \mathbf{S}_2 &\longrightarrow \mathbf{S}_3 + \mathbf{S}_4\ ,\\ \mathbf{R}_2:\ 2\,\mathbf{S}_1 &\longrightarrow \mathbf{S}_1 + \mathbf{S}_5\ , \quad \text{and}\\ \mathbf{R}_3:\ \mathbf{S}_3 &\longrightarrow \mathbf{S}_5\ .\end{aligned} \tag{9.59}$$

The functions $h_{\boldsymbol{\mu}}(\boldsymbol{n})$ are obtained by simple combinatorics

$$\begin{aligned}h_1(\boldsymbol{n}) &= n_1\, n_2\ ,\\ h_2(\boldsymbol{n}) &= n_1\,(n_1 - 1)/2\ , \quad \text{and}\\ h_3(\boldsymbol{n}) &= n_3\ ,\end{aligned}$$

and the matrix $\mathbf{S}$ is of the form

$$\mathbf{S} \;=\; \begin{pmatrix} -1 & -1 & 0 \\ -1 & 0 & 0 \\ +1 & 0 & -1 \\ +1 & 0 & 0 \\ 0 & +1 & +1 \end{pmatrix} \;.$$

It is worth noticing that the functional form of $h_{\boldsymbol{\mu}}$ is determined exclusive by the *reactant side* of $\mathbf{R}_{\boldsymbol{\mu}}$. In particular, has precisely the same form determined by *mass action* in the deterministic kinetic equations with the exception that the particle numbers have to counted exactly in small systems, $n(n-1)$ instead of n^2 for example. The stoichiometric matrix $\mathbf{S}$ refers to the product side of the reaction equations and counts the net production of molecular species per one elementary reaction event: $\nu_{k\boldsymbol{\mu}}$ is the number of molecules $\mathbf{S}_k$ produces by reaction $\mathbf{R}_{\boldsymbol{\mu}}$, these numbers are integers and negative values indicate the number of molecules, which have disappeared during one reaction. In the forthcoming analysis we shall make use of vectors corresponding to individual reactions $\mathbf{R}_{\boldsymbol{\mu}}$: $\boldsymbol{\nu}_{\boldsymbol{\mu}} = (\nu_{1\boldsymbol{\mu}}, \ldots, \nu_{N\boldsymbol{\mu}})'$.

Analogy to deterministic kinetics. It is illustrative to consider now the analogy to conventional chemical kinetics. If we denote the concentration vector of our molecular species $\mathbf{S}_k$ by $\boldsymbol{x} = (x_1, \ldots, x_N)'$ and the flux vector $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_N)'$ the kinetic equation can be expressed by

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{dt}} \;=\; \mathbf{S} \cdot \boldsymbol{\varphi} \;. \tag{9.60}$$

The individual elements of the flux vector in mass action kinetics are

$$\varphi_{\boldsymbol{\mu}} \;=\; k_{\boldsymbol{\mu}} \prod_{k=1}^{n} x_k^{g_{k\boldsymbol{\mu}}} \quad \text{for} \quad g_{1\boldsymbol{\mu}}\,\mathbf{S}_1 \,+\, g_{2\boldsymbol{\mu}}\,\mathbf{S}_2 \,+\, \ldots \,+\, g_{N\boldsymbol{\mu}}\,\mathbf{S}_N \;\longrightarrow$$

wherein the factors $g_{k\boldsymbol{\mu}}$ are the stoichiometric coefficients on the reactant side of the reaction equations. It is sometimes useful to define analogous factors $q_{k\boldsymbol{\mu}}$ for the product side, both classes of factors can be summarized in matrices G and Q and then the stochastic matrix is simply given by the

difference $\mathbf{S} = \mathrm{Q} - \mathrm{G}$. We illustrate by means of the model mechanism (9.59) in our example:

$$\mathrm{Q} - \mathrm{G} = \begin{pmatrix} 0 & +1 & 0 \\ 0 & 0 & 0 \\ +1 & 0 & 0 \\ +1 & 0 & 0 \\ 0 & +1 & +1 \end{pmatrix} - \begin{pmatrix} +1 & +2 & 0 \\ +1 & 0 & 0 \\ 0 & 0 & +1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} -1 & -1 & 0 \\ -1 & 0 & 0 \\ +1 & 0 & -1 \\ +1 & 0 & 0 \\ 0 & +1 & +1 \end{pmatrix} = \mathbf{S}$$

We remark that the entries of G and Q are nonnegative integers by definition. The flux $\boldsymbol{\varphi}$ has the same structure as in the stochastic approach, $\gamma_{\boldsymbol{\mu}}$ corresponds to the kinetic rate parameter or rate constant $k_{\boldsymbol{\mu}}$ and the combinatorial function $h_{\boldsymbol{\mu}}$ and the mass action product are identical apart from the simplifications for large particle numbers.

Occurrence of reactions. The probability of occurrence of reaction events within an infinitesimal time interval dt is cast into three theorems:

Theorem 1. If $\vec{\mathcal{X}}(t) = \boldsymbol{n}$, 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

$$\gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\boldsymbol{n})\ \mathrm{dt} \ + \ o(\,\mathrm{dt})\ ,$$

where $o(\,\mathrm{dt})$ denotes terms that approach zero with dt faster than dt.

Theorem 2. If $\vec{\mathcal{X}}(t) = \boldsymbol{n}$, then the probability that *no* reaction will occur within the time interval $[t, t + \mathrm{dt}[$ is equal to

$$1 \ - \ \sum_{\boldsymbol{\mu}} \gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\boldsymbol{n})\ \mathrm{dt} \ + \ o(\,\mathrm{dt})\ .$$

Theorem 3. The probability of more than one reaction occurring in the system within the time interval $[t, t + \mathrm{dt}[$ is of order $o(\,\mathrm{dt})$.

Proofs for all three theorems are found in [115, pp.420,421].

Based on the three theorems an analytical description of the evolution of the population vector $\vec{\mathcal{X}}(t)$. The initial state of the system at some initial time t_0 is fixed: $\vec{\mathcal{X}}(t_0) = \boldsymbol{n}_0$. Although there is no chance to derive a deterministic equation for the time-evolution, a deterministic function for

the time-evolution of the probability function $P(\boldsymbol{n}, t|\boldsymbol{n}_0, t_0)$ for $t \geq t_0$ will be obtained. We express the probability $P(\boldsymbol{n}, t|\boldsymbol{n}_0, t_0)$ as the sum of the probabilities of several mutually exclusive and and collectively exhaustive *routes* from $\vec{\mathcal{X}}(t_0) = \boldsymbol{n}_0$ to $\vec{\mathcal{X}}(t + \mathrm{dt}) = \boldsymbol{n}$. These routes are distinguished from one another with respect to the event that happened in the time interval $[t, t + \mathrm{dt}[$:

$$\begin{aligned} P(\boldsymbol{n}, t + \mathrm{dt}|\boldsymbol{n}_0, t_0) \;=\; & P(\boldsymbol{n}, t|\boldsymbol{n}_0, t_0) \;\times\; \left(1 - \sum_{\boldsymbol{\mu}=1}^{M} \gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\boldsymbol{n})\,\mathrm{dt} \;+\; o(\,\mathrm{dt})\right) \;+ \\ & + \;\sum_{\boldsymbol{\mu}=1}^{M} P(\boldsymbol{n} - \boldsymbol{\nu}_{\boldsymbol{\mu}}, t|\boldsymbol{n}_0, t_0) \;\times\; \Big(\gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\boldsymbol{n} - \boldsymbol{\nu}_{\boldsymbol{\mu}})\,\mathrm{dt} \;+\; o(\,\mathrm{dt})\Big) \;+ \\ & + \;o(\,\mathrm{dt})\;. \end{aligned} \tag{9.61}$$

The different routes from $\vec{\mathcal{X}}(t_0) = \boldsymbol{n}_0$ to $\vec{\mathcal{X}}(t + \mathrm{dt}) = \boldsymbol{n}$ are obvious from the balance equation (9.61):

(i) One route from $\vec{\mathcal{X}}(t_0) = \boldsymbol{n}_0$ to $\vec{\mathcal{X}}(t + \mathrm{dt}) = \boldsymbol{n}$ 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 $\vec{\mathcal{X}}(t) = \boldsymbol{n}$ was fulfilled at time t. The joint probability for route (i) is therefore *the probability to be in* $\vec{\mathcal{X}}(t) = \boldsymbol{n}$ *conditioned by* $\vec{\mathcal{X}}(t_0) = \boldsymbol{n}_0$ 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 $\boldsymbol{n}_0$ at time t_0 to $\boldsymbol{n}$ at time t and to stay in this state during the next interval dt.

(ii) An alternative route from $\vec{\mathcal{X}}(t_0) = \boldsymbol{n}_0$ to $\vec{\mathcal{X}}(t + \mathrm{dt}) = \boldsymbol{n}$ accounted for by one particular term in sum of terms on the right-hand side of the equation: *An* $\mathbf{R}_{\boldsymbol{\mu}}$ *reaction* is occurring in the time interval $[t, t + \mathrm{dt}[$ and hence $\vec{\mathcal{X}}(t) = \boldsymbol{n} - \boldsymbol{\nu}_{\boldsymbol{\mu}}$ was fulfilled at time t. The joint probability for route (ii) is therefore *the probability to be in* $\vec{\mathcal{X}}(t) = \boldsymbol{n} - \boldsymbol{\nu}_{\boldsymbol{\mu}}$ *conditioned by* $\vec{\mathcal{X}}(t_0) = \boldsymbol{n}_0$ times *the probability that exactly one* $\mathbf{R}_{\boldsymbol{\mu}}$ *reaction has occurred in* $[t, t + \mathrm{dt}[$. In other words, the probability for this route is the probability to go from $\boldsymbol{n}_0$ at time t_0 to $\boldsymbol{n} - \boldsymbol{\nu}_{\boldsymbol{\mu}}$ at time t and to undergo an $\mathbf{R}_{\boldsymbol{\mu}}$ during the next interval dt. Obviously, the same consideration is valid for every elementary reaction and we have M terms of this kind.

(iii) A third possibility – neither *no* reaction nor *exactly one* reaction chosen from the set $\{\mathbf{R}_{\boldsymbol{\mu}}; \boldsymbol{\mu} = 1, \ldots, M\}$ – 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 by theorem 3.

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 a simulation friendly form of the *chemical master equation* is straightforward: $P(\boldsymbol{n}, t|\boldsymbol{n}_0, t_0)$ is subtracted from both sides in Equ. (9.61), 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

$$\begin{aligned}\frac{\partial}{\partial t} P(\boldsymbol{n}, t|\boldsymbol{n}_0, t_0) \;=\; \sum_{\boldsymbol{\mu}=1}^{M} \Big(& \gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\boldsymbol{n} - \boldsymbol{\nu}_{\boldsymbol{\mu}})\, P(\boldsymbol{n} - \boldsymbol{\nu}_{\boldsymbol{\mu}}|\boldsymbol{n}_0, t_0) - \\ & - \gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\boldsymbol{n})\, P(\boldsymbol{n}, t|\boldsymbol{n}_0, t_0) \Big) \;.\end{aligned} \tag{9.62}$$

Initial conditions are required to calculate the time evolution of the probability $P(\boldsymbol{n}, t|\boldsymbol{n}_0, t_0)$ and we can easily express them in the form

$$P(\boldsymbol{n}, t_0|\boldsymbol{n}_0, t_0) \;=\; \begin{cases} 1\,, & \text{if } \boldsymbol{n} = \boldsymbol{n}_0\,, \\ 0\,, & \text{if } \boldsymbol{n} \neq \boldsymbol{n}_0\,, \end{cases} \tag{9.62'}$$

which is precisely the initial condition used in the derivation of equation (9.61). Any sharp probability distribution $P\big(n_k, t_0|n_k^{(0)}, t_0\big) = \delta(n_k - n_k^{(0)})$ is admitted for the molecular particle numbers at t_0. The assumption of extended initial distributions is, of course, also possible but the corresponding master equation becomes more sophisticated.

9.3.3 The simulation algorithm

The chemical master equation (9.62) as derived in the last subsection 9.3.1 was found to be well suited for the derivation of a stochastic simulation algorithm for chemical reactions [112, 113, 117] and it is important to see how the simulation tool fits into the general theoretical framework of the

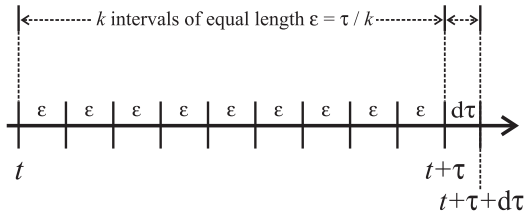


Figure 9.9: **Partitioning of the time interval** $[t, t+\tau+\mathrm{d}\tau[$**.** The entire interval is subdivided into $(k+1)$ nonoverlapping subintervals. The first k intervals are of equal size $\varepsilon = \tau/k$ and the $(k+1)$-th interval is of length $\mathrm{d}\tau$.

chemical master equation. The algorithm is not based on the probability function $P(\boldsymbol{n}, t|\boldsymbol{n}_0, t_0)$ but on another related probability density $p(\tau, \boldsymbol{\mu}|\boldsymbol{n}, t)$, which expresses the probability that given $\vec{\mathcal{X}}(t) = \boldsymbol{n}$ 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.

Considering the theory of random variables, $p(\tau, \boldsymbol{\mu}|\boldsymbol{n}, t)$ 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 $1 \leq \boldsymbol{\mu} \leq M$. In order to derive an explicit formula for the probability density $p(\tau, \boldsymbol{\mu}|\boldsymbol{n}, t)$ we introduce the quantity

$$a(\boldsymbol{n}) \;=\; \sum_{\boldsymbol{\mu}=1}^{M} \gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\boldsymbol{n})$$

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 9.9. With $\vec{\mathcal{X}}(t) = \boldsymbol{n}$ the probability $p(\tau, \boldsymbol{\mu}|\boldsymbol{n}, t)$ 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 theorems 1 and 2 and the multiplication law of probabilities we find

$$p(\tau, \boldsymbol{\mu}|\boldsymbol{n}, t) \;=\; \Big(1 \,-\, a(\boldsymbol{n})\,\varepsilon \,+\, o(\varepsilon)\Big)^{k} \Big(\gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\boldsymbol{n})\,\mathrm{d}\tau \,+\, o(\mathrm{d}\tau)\Big)$$

Dividing both sides by $\mathrm{d}\tau$ and taking the limit $\mathrm{d}\tau \downarrow 0$ yields

$$p(\tau, \boldsymbol{\mu}|\boldsymbol{n}, t) \;=\; \Bigl(1 \,-\, a(\boldsymbol{n})\,\varepsilon \,+\, o(\varepsilon)\Bigr)^k \,\gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\boldsymbol{n})$$

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

$$\begin{aligned}\Bigl(1 \,-\, a(\boldsymbol{n})\,\varepsilon \,+\, o(\varepsilon)\Bigr)^k \;&=\; \left(1 \,-\, \frac{a(\boldsymbol{n})\,k\varepsilon \,+\, k\,o(\varepsilon)}{k}\right)^k \;= \\ &=\; \left(1 \,-\, \frac{a(\boldsymbol{n})\,\tau \,+\, \tau\,o(\varepsilon)/\varepsilon}{k}\right)^k \,,\end{aligned}$$

and take now the limit $k \to \infty$ whereby we make use of the simultaneously occurring convergence $o(\varepsilon)/\varepsilon \downarrow 0$:

$$\lim_{k\to\infty}\Bigl(1 \,-\, a(\boldsymbol{n})\,\varepsilon \,+\, o(\varepsilon)\Bigr)^k \;=\; \lim_{k\to\infty}\left(1 - \frac{a(\boldsymbol{n})\,\tau}{k}\right)^k \;=\; e^{-a(\boldsymbol{n})\,\tau}\;.$$

By substituting this result into the initial equation for the probability density of the occurrence of a reaction we find

$$p(\tau, \boldsymbol{\mu}|\boldsymbol{n}, t) \;=\; a(\boldsymbol{n})\, e^{-a(\boldsymbol{n})\,\tau}\;\frac{\gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\boldsymbol{n})}{a(\boldsymbol{n})} \;=\; \gamma_{\boldsymbol{\mu}}\, h_{\boldsymbol{\mu}}(\boldsymbol{n})\; e^{-\sum_{\nu=1}^{M}\gamma_\nu h_\nu(\boldsymbol{n})\,\tau}. \quad (9.63)$$

Equ. (9.63) provides the mathematical basis for the stochastic simulation algorithm. Given $\vec{\mathcal{X}}(t) = \boldsymbol{n}$, 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 r_1 and r_2.

9.3.4 Implementation of the simulation algorithm

Equ. (9.63) is implemented now for computer simulation and we inspect the probability densities of the two unit-interval uniform random variables r_1 and r_2 in order to find the conditions to be imposed of a statistically exact

sample pair $(\tau, \boldsymbol{\mu})$: r_1 has an exponential density function with the decay constant $a(\boldsymbol{n})$,

$$\tau \;=\; \frac{1}{a(\boldsymbol{n})}\;\ln\bigl(1\;/\;r_1\bigr)\;, \tag{9.64a}$$

and taking m to be the *smallest* integer which fulfils

$$\boldsymbol{\mu} \;=\; \inf\left\{m \;\Bigg|\; \sum_{\mu=1}^{m} c_\mu\, h_\mu(\boldsymbol{n}) \;>\; a(\boldsymbol{n})\; r_2\right\}\;. \tag{9.64b}$$

After the values for τ and $\boldsymbol{\mu}$ have been determined accordingly the action *advance the state vector* $\vec{\mathcal{X}}(t)$ *of the system* is taking place:

$$\vec{\mathcal{X}}(t) \;=\; \boldsymbol{n} \quad\longrightarrow\quad \vec{\mathcal{X}}(t+\tau) = \boldsymbol{n} + \boldsymbol{\nu_\mu}\;.$$

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 r_1 and r_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 $\vec{\mathcal{X}}(t) \;=\; \boldsymbol{n}$ throughout the entire finite time interval $[t, t + \mathrm{d}\tau[$ and then changes abruptly to $\vec{\mathcal{X}}(t+\tau) = \boldsymbol{n} + \boldsymbol{\nu_\mu}$ 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.

Table 9.1: **The combinatorial functions $h_{\boldsymbol{\mu}}(\boldsymbol{n})$ 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 1,2 and 3 mean that one, two or three molecules are involved in the elementary step, respectively.

No.	Reaction	Order	$h_{\boldsymbol{\mu}}(\boldsymbol{n})$
1	$* \longrightarrow$ products	0	1
2	$\mathsf{A} \longrightarrow$ products	1	n_{A}
3	$\mathsf{A}+\mathsf{B} \longrightarrow$ products	2	$n_{\mathsf{A}}n_{\mathsf{B}}$
4	$2\,\mathsf{A} \longrightarrow$ products	2	$n_{\mathsf{A}}(n_{\mathsf{A}}-1)/2$
5	$\mathsf{A}+\mathsf{B}+\mathsf{C} \longrightarrow$ products	3	$n_{\mathsf{A}}n_{\mathsf{B}}n_{\mathsf{C}}$
6	$2\,\mathsf{A}+\mathsf{B} \longrightarrow$ products	3	$n_{\mathsf{A}}(n_{\mathsf{A}}-1)n_{\mathsf{B}}/2$
7	$3\,\mathsf{A} \longrightarrow$ products	3	$n_{\mathsf{A}}(n_{\mathsf{A}}-1)(n_{\mathsf{A}}-2)/6$

Structure of the algorithm. The time evolution of the population in described by the vector $\vec{\mathcal{X}}(t) = \boldsymbol{n}(t)$, which is updated after every individual reaction event. Reactions are chosen from the set $\mathcal{R} = \{\mathbf{R}_{\boldsymbol{\mu}}; \boldsymbol{\mu} = 1, \ldots, M\}$, which is defined by the reaction mechanism under consideration. They are classified according to the criteria listed in table 9.1. The reaction probabilities corresponding to the reaction rates of deterministic kinetics are contained in a vector $\boldsymbol{a}(\boldsymbol{n}) = \big(c_1 h_1(\boldsymbol{n}), \ldots, c_M h_M(\boldsymbol{n})\big)'$, which is also updated after every individual reaction event. Updating is performed according to the stoichiometric vectors $\boldsymbol{\nu}_{\boldsymbol{\mu}}$ of the individual reactions $\mathbf{R}_{\boldsymbol{\mu}}$, which represent columns of the stoichiometric matrix $\mathbf{S}$. We repeat that the combinatorial functions $h_{\boldsymbol{\mu}}(\boldsymbol{n})$ are determined exclusively by the reactant side of the reaction equation whereas the stoichiometric vectors $\boldsymbol{\nu}_{\boldsymbol{\mu}}$ represent the net production, (*products*)$-$(*reactants*).

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 $\mathbf{S}_k$ – are stored, the values for the M parameters of the reactions $\mathbf{R}_{\boldsymbol{\mu}}$, $c_1, \ldots, c_M$, are stored, and the combinatorial expressions are incorporated as factors for the calculation of the reaction rate vector $\boldsymbol{a}(\boldsymbol{n})$ according to table 9.1 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 $\boldsymbol{n}$, $t \to t + \tau$ and $\boldsymbol{n} \to \boldsymbol{n} + \boldsymbol{\nu}_{\boldsymbol{\mu}}$, then all changes are incorporated in a recalculation of the reaction rate vector $\boldsymbol{a}$.

(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 $\boldsymbol{n}(t)$ 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 $h_{\boldsymbol{\mu}} = 0 \; \forall \; \boldsymbol{\mu} = 1, \ldots, M$, 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 $\boldsymbol{n}$ 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:

$$0 \leq \alpha < \beta \leq 1 \qquad \Longrightarrow \qquad P(\alpha \leq r \leq \beta) \;=\; \beta - \alpha \;.$$

With this prerequisite we discuss now two methods which use two output values r of the pseudorandom number generator to generate a random pair $(\tau, \boldsymbol{\mu})$ with the prescribed probability density function $P(\tau, \boldsymbol{\mu})$. Two implementations are frequently used:

(i) 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 $\mathbf{R}_{\boldsymbol{\mu}}$:

$$P_1(\tau) \;=\; \sum_{\boldsymbol{\mu}=1}^{M} P(\tau, \boldsymbol{\mu}) \;. \tag{9.65}$$

Combining the last two equations we obtain for $P_2(\boldsymbol{\mu}|\tau)$

$$P_2(\boldsymbol{\mu}|\tau) \;=\; P(\tau, \boldsymbol{\mu}) \;/\; \sum_{\boldsymbol{\nu}}^{M} P(\tau, \boldsymbol{\nu}) \tag{9.66}$$

Equations (9.65) and (9.66) express the two one-variable density functions in terms of the original two-variable density function $P(\tau, \boldsymbol{\mu})$. From equation (9.63) we substitute into $P(\tau, \boldsymbol{\mu}) = p(\tau, \boldsymbol{\mu}|\boldsymbol{n}, t)$ through simplifying the notation by using

$$a_{\boldsymbol{\mu}} \;\equiv\; \gamma_{\boldsymbol{\mu}} h_{\boldsymbol{\mu}}(\boldsymbol{n}) \;\text{ and }\; a \;=\; \sum_{\boldsymbol{\mu}=1}^{M} a_{\boldsymbol{\mu}} \;\equiv\; \sum_{\boldsymbol{\mu}=1}^{M} \gamma_{\boldsymbol{\mu}} h_{\boldsymbol{\mu}}(\boldsymbol{n})$$

and find

$$\begin{aligned} P_1(\tau) &= a\,\exp(-a\,\tau)\,, \quad 0 \leq \tau < \infty \ \text{ and} \\ P_2(\boldsymbol{\mu}|\tau) &= P_2(\boldsymbol{\mu}) = a_{\boldsymbol{\mu}} \,/\, a\,, \quad \boldsymbol{\mu} = 1, \ldots, M\ . \end{aligned} \tag{9.67}$$

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:

$$\int_0^\infty P_1(\tau)\,\mathrm{d}\tau = \int_0^\infty a\,e^{-a\,\tau}\,\mathrm{d}\tau = 1 \ \text{ and } \ \sum_{\boldsymbol{\mu}=1}^{M} P_2(\boldsymbol{\mu}) = \sum_{\boldsymbol{\mu}=1}^{M} \frac{a_{\boldsymbol{\mu}}}{a} = 1\ .$$

Thus, in the *direct* method a random value τ is created from a random number on the unit interval, r_1, and the distribution $P_1(\tau)$ by taking

$$\tau = -\,\frac{\ln r_1}{a}\ . \tag{9.68}$$

The second task is to generate a random integer $\hat{\boldsymbol{\mu}}$ 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, r_2, will be drawn from the unit interval and then $\hat{\boldsymbol{\mu}}$ is taken to be the integer that fulfils

$$\sum_{\boldsymbol{\nu}=1}^{\boldsymbol{\mu}-1} a_{\boldsymbol{\nu}} < r_2\,a \leq \sum_{\boldsymbol{\nu}=1}^{\boldsymbol{\mu}} a_{\boldsymbol{\nu}}\ . \tag{9.69}$$

The values a_1, a_2, ... , are cumulatively added in sequence until their sum is observed to be equal or to exceed $r_2 a$ and then $\hat{\boldsymbol{\mu}}$ is set equal to the index of the last $a_{\boldsymbol{\nu}}$ term that had been added. Rigorous justifications for equations (9.68) and (9.69) are found in [112, 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.

(ii) 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 $a_{\boldsymbol{\nu}} \equiv \gamma_{\boldsymbol{\nu}} h_{\boldsymbol{\nu}}(\boldsymbol{n})$ it is straightforward to derive

$$P_{\boldsymbol{\nu}}(\tau)\,\mathrm{d}\tau \;=\; a_{\boldsymbol{\nu}}\,\exp(-a_{\boldsymbol{\nu}}\,\tau)\,\mathrm{d}\tau \tag{9.70}$$

from (9.56) and (9.57). 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 $r_{\boldsymbol{\nu}}$ from the unit interval and compute

$$\tau_{\boldsymbol{\nu}} \;=\; -\frac{\ln r_{\boldsymbol{\nu}}}{a_{\boldsymbol{\nu}}}\,,\; \boldsymbol{\nu} = 1,\ldots,M\;. \tag{9.71}$$

From these M *tentative next* reactions the one, which occurs first, is chosen to be the *actual next* reactions:

$$\begin{aligned} \tau \;&=\; \text{smallest } \tau_{\boldsymbol{\nu}} \;\text{ for all }\; \boldsymbol{\nu} = 1,\ldots,M\;, \\ \boldsymbol{\mu} \;&=\; \boldsymbol{\nu} \text{ for which } \tau_{\boldsymbol{\nu}} \text{ is smallest}\,. \end{aligned} \tag{9.72}$$

Daniel Gillespie [112, 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 (9.63).

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, $\boldsymbol{n}$, because the results of the first reaction are not incorporated into the combinatorial terms $h_{\boldsymbol{\mu}}(\boldsymbol{n})$. 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 ion the first reaction.

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 $M \geq 3$ 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 $M \geq 3$ and in particular for large M the direct method is probably the method of choice for the Monte Carlo step.

An early computer code of the simple version of the algorithm described – still in *FORTRAN* – is found in [112]. Meanwhile many attempts were made in order to speed-up computations and allow for simulation of stiff systems (see e.g. [33]. A recent review of the simulation methods also contains a discussion of various improvements of the original code [117].

10. Stochasticity in evolution

The population aspect is particularly important in evolution and accordingly we stress it again here. A population vector

$$\mathbf{\Pi}(t) = \big(N_1(t), N_2(t), \ldots, N_n(t)\big) \quad \text{with} \quad N_k \in \mathbb{N}^0\,,\ t \in \mathbb{R}^1_+\,,$$

counts the numbers of individuals for the different *species*[1] $\mathbf{X}_k$ as a function of time $N_k(t)$. Implicitly this kind of formulating the problem states already that time will be considered as a continuous variable and *counting* says that the numbers of individuals are considered to vary in discrete steps. The basic assumptions thus are the same as in the applications of master equations to chemical reaction kinetics (section 9.1). There is one major difference between the molecular approach based on elementary reactions and macroscopic modeling often used in biology. The objects are no longer single molecules or atoms but modules commonly consisting of a large number of atoms or individual cells or organisms: Elementary step dynamics fulfils several conservation relations like conservation of mass and conservation of the number of atoms of every chemical element (unless nuclear reactions are admitted), and the laws of thermodynamics provide additional restrictions. In the macroscopic models these relations are not violated, of course, but they are hidden in complex networks of interactions, which appear in the model only after averaging on several hierarchical levels. For example, conservation of mass and energy are encapsulated and obscured in the carrying capacity K of the ecosystem as modeled by the Verhulst equation (1.7). As a consequence the numbers of individuals may change in biological models, $N_k(t) \to N_k(t + \Delta t) \pm 1$, without a compensation in another variable. The analogue in chemistry is *buffering* where a large molecular reservoir remains

[1] In this chapter we use species in the sense of *molecular species* and indicate by *biological species* the genuine notion of species in biology.

practically unchanged when a single molecule is added or subtracted (see section 9.2, irreversible addition reaction). Changes ± 1 in the numbers of individuals imply that the time interval considered is sufficiently short that multiple events can be excluded. In biology we can understand the flow reactor (section 4.2) as an idealized version of an ecosystem. Then the biological analogues to influx (4.1a) and outflux (4.1d) are migration, immigration and emigration, respectively.

A stochastic process on the population level is a recording of ordered successive events at times T_i:

$$T_0 < T_1 < T_2 < \ldots < T_j < T_k \ldots ,$$

along a continuous time axis t.[2] A birth or death event at some time $t = T_r$, for example, creates, $\mathbf{X}_j \rightarrow 2\mathbf{X}_j$, or consumes one individual, $\mathbf{X}_j \rightarrow \oslash$, and the population changes accordingly:

$$\mathbf{\Pi} = \begin{cases} \big(\ldots, N_j(t) = N_j(T_{r-1}), N_k(t) = N_k(T_{r-1}), \ldots\big) & \text{if } \; T_{r-1} \leq t < T_r \\ \big(\ldots, N_j(t) = N_j(T_{r-1}) \pm 1, N_k(t) = N_k(T_{r-1}), \ldots\big) & \text{if } \; T_r \leq t < T_{r+1} \end{cases} .$$

This formulation of a biological birth or death events reflects the previously mentioned convention in probability theory: Right-hand continuity is assumed for steps in stochastic processes (see Fig. 8.6).

Compared to stochasticity in chemistry stochastic phenomena in biology are not only more important but also much harder to control. The major sources of the problem are small population numbers and the lack of sufficiently simple references systems that are accessible to experimental studies. In biology we are regularly encountering reaction mechanisms that lead to enhancement of fluctuations at non-equilibrium conditions and biology in essence is dealing with processes and stationary states far away from equilibrium whereas in chemistry autocatalysis in non-equilibrium systems became an object of general interest and intensive investigation not before some forty years ago. We start therefore with the analysis of simple autocatalysis modeled by means of a simple birth-and-death process. Then we present an

[2]The application of discretized time in evolution – mimicking synchronized generations, for example – is straightforward but we shall mention only briefly here because we focus on continuous time birth-and-death processes and master equations.

overview over solvable birth-and-death processes (section 10.1) and discuss the role of boundaries in form of different barriers (section 10.1.2).

10.1 Autocatalysis, replication, and extinction

In the previous chapter we analyzed already bimolecular reactions, the addition and the dimerization reaction, which gave rise to perfectly *normal* behavior although the analysis was shown to be quite sophisticated (section 9.2). The nonlinearity manifested itself only in task to find solutions and did not change effectively the qualitative behavior of the reaction systems, for example the $\sqrt{N}$-law for the fluctuations in the stationary states retained its validity. As an exactly solvable example we shall study first a simple reaction mechanism consisting of two elementary steps, autocatalytic replication and extinction. In this case the $\sqrt{N}$-law is not valid and fluctuations do not settle down to some value which is proportional to the square root of the size of the system but grow in time without limit as we saw in case of the Wiener process (8.2.2).

10.1.1 Autocatalytic growth and death

Reproduction of individuals is modeled by a simple duplication mechanism and death is represented by first order decay. In the language of chemical kinetics these two steps are:

$$\mathbf{A} + \mathbf{X} \xrightarrow{\lambda} 2\,\mathbf{X}\,, \tag{10.1a}$$

$$\mathbf{X} \xrightarrow{\mu} \oslash\,. \tag{10.1b}$$

The rate parameters for reproduction and extinction are denoted by λ and μ, respectively.[3] The material required for reproduction is assumed to be replenished as it is consumed and hence the amount of $\mathbf{A}$ available is constant and assumed to be included in the birth parameter: $\lambda = f \cdot [\mathbf{A}]$. Degradation products of $\mathbf{X}$ do not enter the kinetic equation because reaction (10.1b) is

[3]Reproduction is to be understood a asexual reproduction here. Sexual reproduction, of course, requires two partners and gives rise to a process of order two (table 9.1).

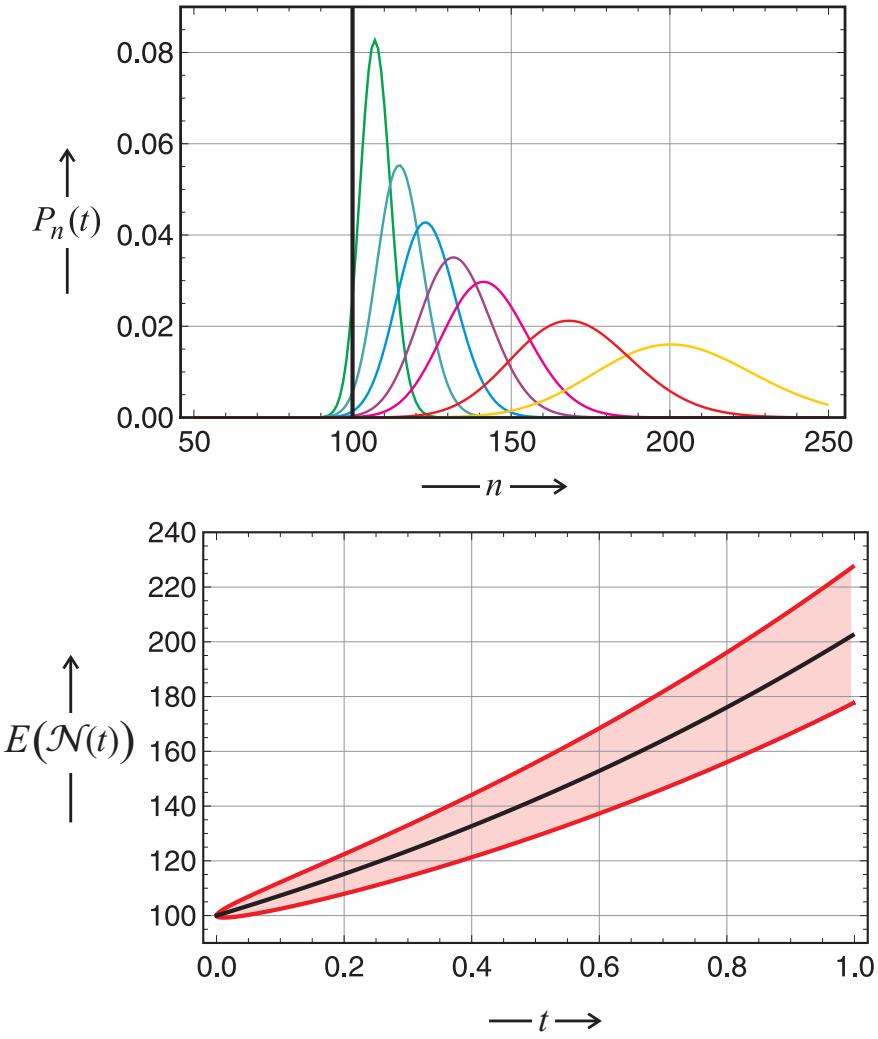


Figure 10.1: **A growing linear birth-and-death process.**The two-step reaction mechanism of the process is ($\mathbf{X} \to 2\mathbf{X}$, $\mathbf{X} \to \varnothing$) with rate parameters λ and μ, respectively. The upper part shows the evolution of the probability density, $P_n(t) = \text{Prob}\,\mathcal{X}(t) = n$. The initially infinitely sharp density, $P(n,0) = \delta(n,n_0)$ becomes broader with time and flattens as the variance increases with time. In the lower part we show the expectation value $E\big(\mathcal{N}(t)\big)$ in the confidence interval $E \pm \sigma$. Parameters used: $n_0 = 100$, $\lambda = \sqrt{2}$, and $\mu = 1/\sqrt{2}$; sampling times (upper part): $t = 0$ (black), 0.1 (green), 0.2 (turquoise), 0.3 (blue), 0.4 (violet), 0.5 (magenta), 0.75 red), and 1.0 (yellow).

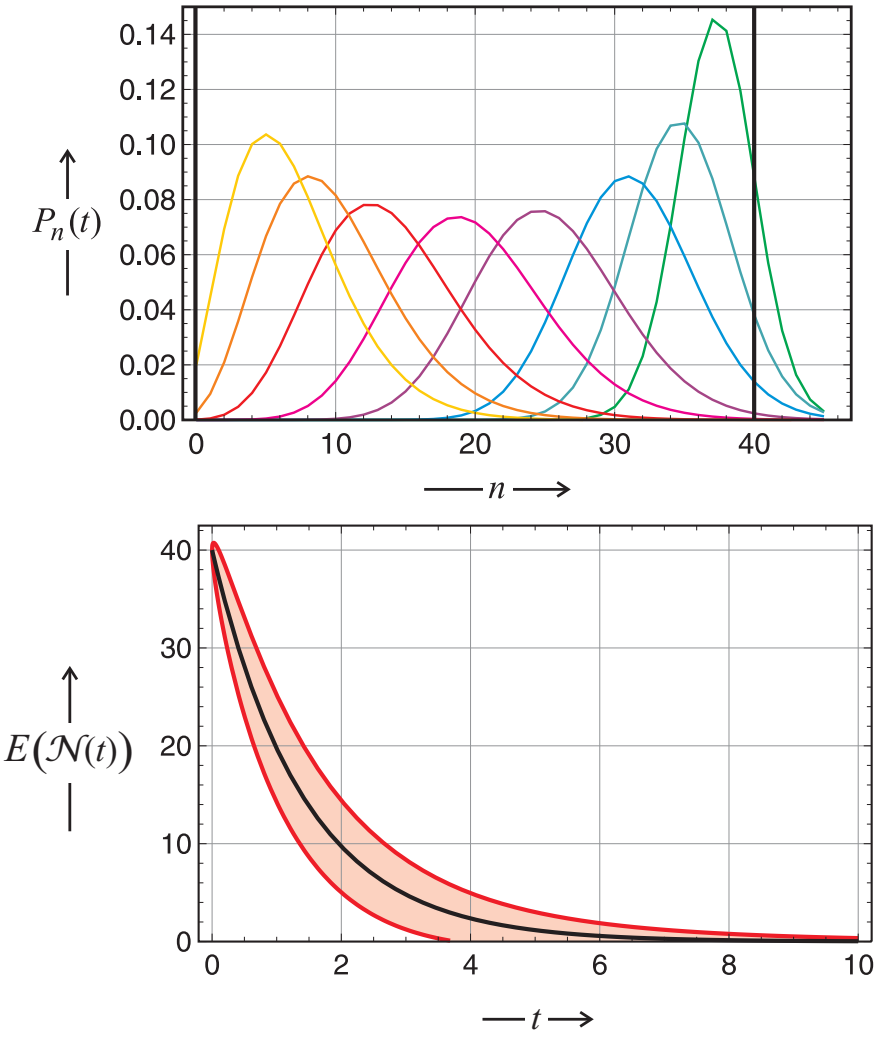


Figure 10.2: **A decaying linear birth-and-death process.** The two-step reaction mechanism of the process is ($\mathbf{X} \to 2\mathbf{X}$, $\mathbf{X} \to \varnothing$) with rate parameters λ and μ, respectively. The upper part shows the evolution of the probability density, $P_n(t) = \text{Prob}\,\mathcal{X}(t) = n$. The initially infinitely sharp density, $P(n,0) = \delta(n,n_0)$ becomes broader with time and flattens as the variance increases but then sharpens again as process approaches the absorbing barrier at $n = 0$. In the lower part we show the expectation value $E\big(\mathcal{N}(t)\big)$ in the confidence interval $E \pm \sigma$. Parameters used: $n_0 = 40$, $\lambda = 1/\sqrt{2}$, and $\mu = \sqrt{2}$; sampling times (upper part): $t = 0$ (black), 0.1 (green), 0.2 (turquoise), 0.35 (blue), 0.65 (violet), 1.0 (magenta), 1.5 red), 2.0 (orange), 2.5 (yellow), and $\lim_{t\to\infty}$ (black).

irreversible. The stochastic process corresponding to equations (10.1) belongs to the class of linear birth-and-death processes with $w^+(n) = \lambda \cdot n$ and $w^-(n) = \mu \cdot n$.[4] The master equation is of the form,

$$\frac{\partial P_n(t)}{\partial t} \;=\; \lambda\,(n-1)\,P_{n-1}(t) \;+\; \mu\,(n+1)\,P_{n+1}(t) \;-\; (\lambda+\mu)\,n\,P_n(t)\;, \quad (10.2)$$

and after introduction of the probability generating function $g(s,t)$ gives rise to the PDE

$$\frac{\partial g(s,t)}{\partial t} \;-\; (s-1)\,(\lambda s-\mu)\,\frac{\partial g(s,t)}{\partial s} \;=\; 0\;. \quad (10.3)$$

Solution of this PDE yields different results for equal or different replication and extinction rate coefficients, $\lambda \neq \mu$ and $\lambda = \mu$, respectively. In the first case we substitute $\gamma = \lambda/\mu$ $(\neq 1)$ and $\eta(t) = \exp\bigl((\lambda-\mu)t\bigr)$, and find:

$$\begin{aligned} g(s,t) \;&=\; \left\{\frac{\bigl(\eta(t)-1\bigr) \;+\; \bigl(\gamma-\eta(t)\bigr)\,s}{\bigl(\gamma\eta(t)-1\bigr) \;+\; \gamma\,\bigl(1-\eta(t)\bigr)\,s}\right\}^{n_0} \qquad \text{and} \\ P_n(t) \;&=\; \gamma^n \sum_{m=0}^{\min(n,n_0)} (-1)^m \binom{n_0+n-m-1}{n-m}\binom{n_0}{m} \times \\ &\qquad \times \left(\frac{1-\eta(t)}{1-\gamma\eta(t)}\right)^{n_0+n-m} \left(\frac{\gamma-\eta(t)}{\gamma\bigl(1-\eta(t)\bigr)}\right)^{m}\;. \end{aligned} \quad (10.4)$$

In the derivation of the expression for the probability distributions we expanded enumerator and denominator of the expression in the generating function $g(s,t)$, by using expressions for the sums $(1+s)^n = \sum_{k=0}^{n}\binom{n}{k}s^k$ and $(1+s)^{-n} = 1+\sum_{k=1}^{\infty}(-1)^k\,\frac{n(n+1)\ldots(n+k-1)}{k!}\,s^k$, multiply, order terms with respect to powers of s, and compare with the expansion of the generating function, $g(s,t) = \sum_{n=0}^{\infty} P_n(t)\,s^n$.

[4]Here we use the symbols commonly applied in biology: $\lambda_{(n)}$ for birth, $\mu_{(n)}$ for death, ν for immigration, and ρ for emigration (tables 10.1 and 10.2). These notions were created especially for application to biological problems, in particular for problems in theoretical ecology. Other notions and symbols are common in chemistry: A birth corresponds to the production of a molecule, $f \equiv \lambda$, a death to its decomposition or degradation through a chemical reaction, $d \equiv \mu$. Influx and outflux are the proper notions for immigration and emigration.

Computations of expectation value and variance are straightforward:

$$\begin{aligned} E\big(\mathcal{N}_X(t)\big) &= n_0\, e^{(\lambda-\mu)\,t} \quad \text{and} \\ \sigma^2\big(\mathcal{N}_X(t)\big) &= n_0\, \frac{\lambda+\mu}{\lambda-\mu}\, e^{(\lambda-\mu)\,t}\, \big(e^{(\lambda-\mu)\,t}-1\big) \end{aligned} \tag{10.5}$$

Illustrative examples of linear birth-and-death processes with growing ($\lambda > \mu$) and decaying ($\lambda < \mu$) populations are shown in figures 10.1 and 10.2, respectively.

In the degenerate case of *neutrality* with respect to growth, $\mu = \lambda$, the same procedure yields:

$$g(s,t) = \left(\frac{\lambda t + (1-\lambda t)\, s}{1+\lambda t + \lambda t\, s}\right)^{n_0}, \tag{10.6a}$$

$$P_n(t) = \left(\frac{\lambda t}{1+\lambda t}\right)^{n_0+n} \sum_{m=0}^{\min(n,n_0)} \binom{n_0+n-m-1}{n-m} \binom{n_0}{m} \left(\frac{1-\lambda^2 t^2}{\lambda^2 t^2}\right)^m, \tag{10.6b}$$

$$E\big(\mathcal{N}_X(t)\big) = n_0\,, \text{ and} \tag{10.6c}$$

$$\sigma^2\big(\mathcal{N}_X(t)\big) = 2\, n_0\, \lambda t\ . \tag{10.6d}$$

Comparison of the last two expressions shows the inherent instability of this reaction system. The expectation value is constant whereas the fluctuations increase with time. The degenerate birth-and-death process is illustrated in Fig. 10.3. The case of steadily increasing fluctuations is in contrast to an equilibrium situation where both, expectation value and variance approach constant values. Recalling the Ehrenfest urn game, where fluctuations were negatively correlated with the deviation from equilibrium, we have here two uncorrelated processes, replication and extinction. The number od individuals n fulfils a kind of random walk on the natural numbers, and indeed in case of the random walk (see Equ. (8.77) in subsection 8.2.1 we had also obtained a constant expectation value $E = n_0$ and a variance that increases linearly with time, $\sigma^2(t) = 2\vartheta(t-t_0)$).

A constant expectation value accompanied by a variance that increases with time has an easy to recognize consequence: At some critical time above which the standard deviation exceeds the expectation, $t_{\rm cr} = n_0 \,/\, (2\lambda)$. From

this instant on predictions on the evolution of the system based on the expectation value become obsolete. Then we have to rely on individual probabilities or other quantities. Useful in this context is the probability of extinction of all individuals, which can be readily computed:

$$P_0(t) \;=\; \left(\frac{\lambda t}{1+\lambda t}\right)^{n_0} \quad . \tag{10.7}$$

Provided we wait long enough, the system will die out with probability one, since we have $\lim_{t\to\infty} P_0(t) = 1$. This seems to be a contradiction to the constant expectation value. As a matter of fact it is not: In almost all individual runs the system will go extinct, but there are very few cases of probability measure zero where the number of individuals grows to infinity for $t \to \infty$. These rare cases are responsible for the finite expectation value.

Equ. (10.7) can be used to derive a simple model for *random selection* [254]. We assume a population of n different species

$$\mathbf{A} \;+\; \mathbf{X}_j \quad \xrightarrow{\;\lambda\;} \quad 2\,\mathbf{X}_j \;, \;\; j = 1, \ldots, n \;, \tag{10.1a'}$$

$$\mathbf{X}_j \quad \xrightarrow{\;\mu\;} \quad \oslash \;, \qquad j = 1, \ldots, n \;. \tag{10.1b'}$$

The probability joint distribution of the population is described by

$$P_{x_1 \ldots x_n} \;=\; P\big(\mathcal{X}_1(t) = x_1, \ldots, \mathcal{X}_n(t) = x_n\big) \;=\; P_{x_1}^{(1)} \cdot \ldots \cdot P_{x_n}^{(n)} \;, \tag{10.8}$$

wherein all probability distribution for individual species are given by Equ. (10.6b) and independence of individual birth events as well as death events allows for the simple product expression. In the spirit of Motoo Kimura's neutral theory of evolution [174] all birth and all death parameters are assumed to be equal, $\lambda_j = \lambda$ and $\mu_j = \mu$ for all $j = 1, \ldots, n$. For convenience we assume that every species is initially present in a single copy: $P_{n_j}(0) = \delta_{n_j,1}$. We introduce a new random variable that has the nature of a first passage time: $\mathcal{T}_k$ is the time up to the extinction of $n - k$ species and characterize it as *sequential extinction time*. Accordingly, n species are present in the population between $\mathcal{T}_n$, which fulfils $\mathcal{T}_n \equiv 0$ by definition, and $\mathcal{T}_{n-1}$, $n - 1$ species between $\mathcal{T}_{n-1}$ and $\mathcal{T}_{n-2}$, and eventually a single species between $\mathcal{T}_1$

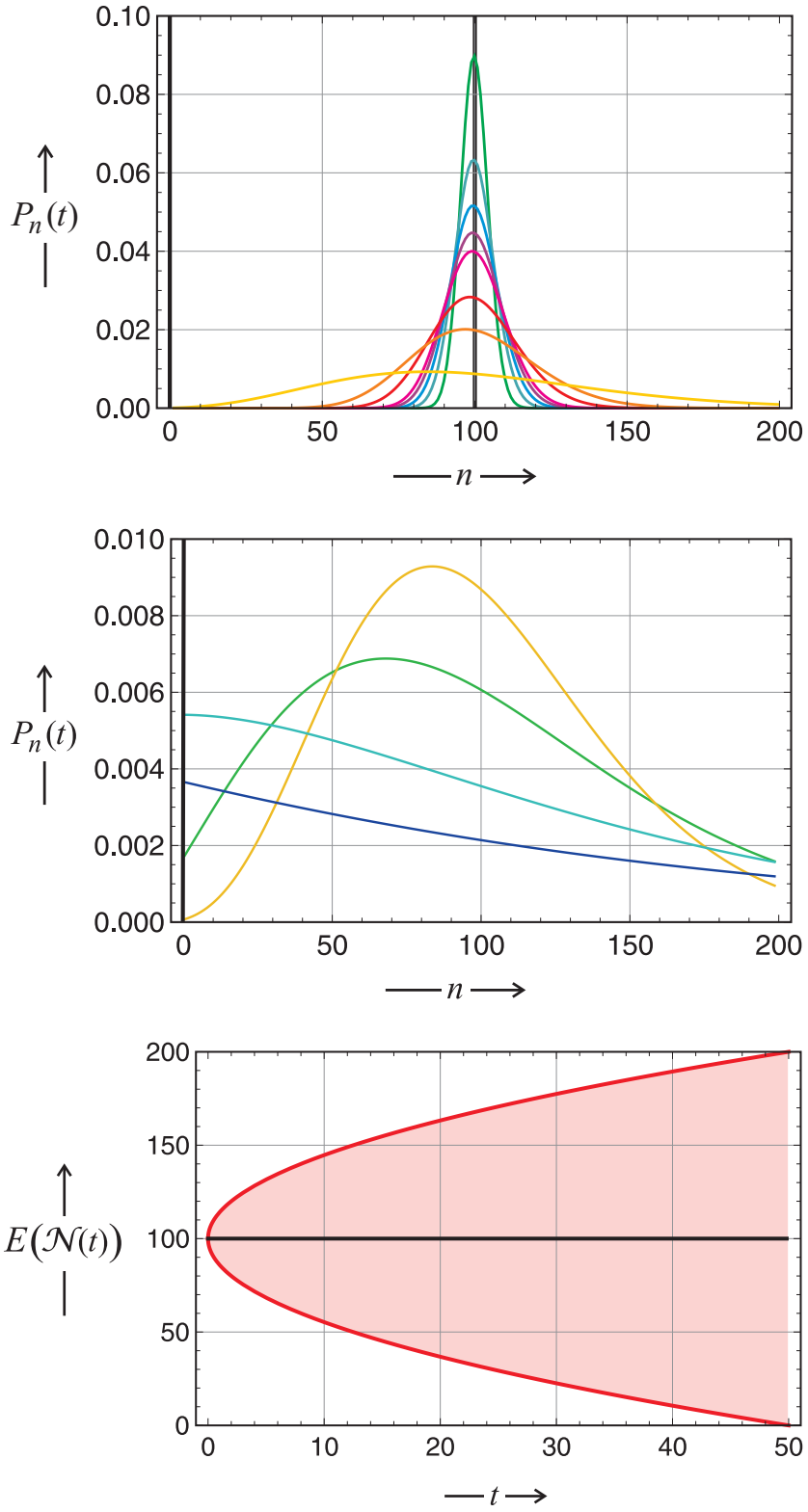


Figure 10.3: Continued on next page.

Figure 10.3: **Probability density of a linear birth-and-death with equal birth and death rate.** The two-step reaction mechanism is: ($\mathbf{X} \rightarrow 2\mathbf{X}$, $\mathbf{X} \rightarrow \oslash$) with rate parameters $\lambda = \mu$. The upper and the middle part show the evolution of the probability density, $P_n(t) = \text{Prob}\big(\mathcal{X}(t) = n\big)$. The initially infinitely sharp density, $P(n,0) = \delta(n,n_0)$ becomes broader with time and flattens as the variance increases but then sharpens again as the process approaches the absorbing barrier at $n = 0$. In the lower part, we show the expectation value $E\big(\mathcal{N}(t)\big)$ in the confidence interval $E \pm \sigma$. The variance increases linearly with time and at $t = n_0/(2\lambda) = 50$ the standard deviation is as large as the expectation value. Parameters used: $n_0 = 100$, $\lambda = 1$; sampling times, upper part: $t = 0$ (black), 0.1 (green), 0.2 (turquoise), 0.3 (blue), 0.4 (violet), 0.49999 (magenta), 0.99999 red), 2.0 (orange), 10 (yellow), and middle part: $t = 10$ (yellow), 20 (green), 50 (cyan), 100 (blue), and $\lim_{t\to\infty}$ (black).

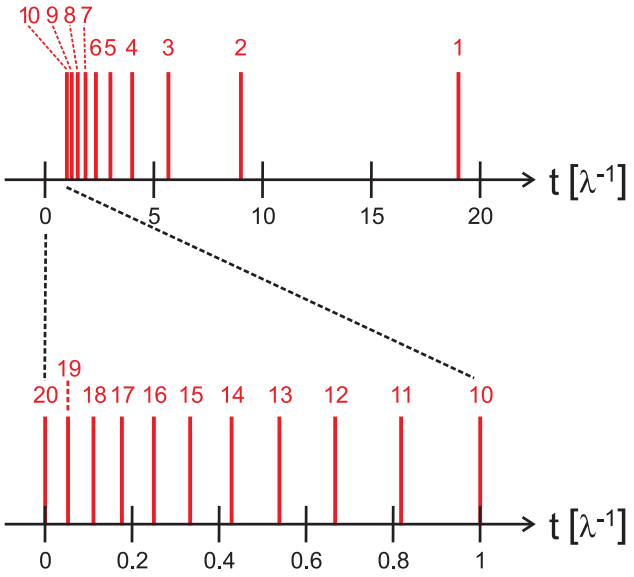


Figure 10.4: **The distribution of sequential extinction times $\mathcal{T}_k$.** Shown are the expectation values $E(\mathcal{T}_k)$ for $n = 20$ according to equation(10.10). Since $E(\mathcal{T}_0)$ diverges, $\mathcal{T}_1$ is the extinction that appears on the average at a finite value. A single species is present above $\mathcal{T}_1$ and random selection has occurred in the population.

and $\mathcal{T}_0$, which is the moment of extinction of the entire population. After $\mathcal{T}_0$ no individual of type $\mathbf{X}$ exists any more.

Next we consider the probability distribution of the sequential extinction times

$$H_k(t) \;=\; P(\mathcal{T}_k < t) \;. \tag{10.9}$$

The probability of extinction of the population is readily calculated: Since individual reproduction and extinction events are independent we find

$$H_0 \;=\; P_{0,\ldots,0} \;=\; P_0^{(1)} \cdot \ldots \cdot P_0^{(n)} \;=\; \left(\frac{\lambda t}{1+\lambda t}\right)^n \;.$$

The event $\mathcal{T}_1 < t$ can happen in several ways: Either $\mathbf{X}_1$ is present and all other species have become extinct already, or only $\mathbf{X}_2$ is present, or only $\mathbf{X}_3$, and so on, but $\mathcal{T}_1 < t$ is also fulfilled if the whole population has died out:

$$H_1 \;=\; P_{x_1\neq 0,0,\ldots,0} \,+\, P_{0,x_2\neq 0,\ldots,0} \,+\, P_{0,0,\ldots,x_n\neq 0} \,+\, H_0 \;.$$

The probability that a given species has not yet disappeared is obtained by exclusion since existence and nonexistence are complementary,

$$P_{x\neq 0} \;=\; 1 \,-\, P_0 \;=\; 1 \,-\, \frac{\lambda t}{1+\lambda t} \;=\; \frac{1}{1+\lambda t} \;,$$

which yields the expression for the presence of a single species

$$H_1(t) \;=\; (n+\lambda t)\frac{(\lambda t)^{n-1}}{(1+\lambda t)^n} \;,$$

and by similar arguments a recursion formula is found for the extinction probabilities with higher indices

$$H_k(t) \;=\; \binom{n}{k} \frac{(\lambda t)^{n-k}}{(1+\lambda t)^n} \,+\, H_{k-1}(t) \;,$$

that eventually leads to the expression

$$H_k(t) \;=\; \sum_{j=0}^{k} \binom{n}{j} \frac{(\lambda t)^{n-j}}{(1+\lambda t)^n} \;.$$

The moments of the sequential extinction times are computed straightforwardly by means of a handy trick: H_k is partitioned into terms for the

individual powers of λt, $H_k(t) = \sum_{j=0}^{k} h_j(t)$ and then differentiated with respect to time t

$$h_j(t) = \binom{n}{j} \frac{(\lambda t)^{n-k}}{(1+\lambda t)^n} ,$$

$$\frac{\mathrm{d}h_j(t)}{\mathrm{dt}} = h'_j = \frac{\lambda}{(1+\lambda t)^{n+1}} \left(\binom{n}{j} (n-j)(\lambda t)^{n-j-1} - \binom{n}{j} j (\lambda t)^{n-j} \right) .$$

The summation of the derivatives is simple because $h'_k + h'_{k-1} + \ldots + h'_0$ is a telescopic sum and we find

$$\frac{\mathrm{d}H_k(t)}{\mathrm{dt}} = \binom{n}{k} (n-k)\, \lambda^{n-k} \frac{t^{n-k-1}}{(1+\lambda t)^{n+1}} .$$

Making use of the definite integral [123, p.338]

$$\int_0^\infty \frac{t^{n-k}}{(1+\lambda t)^{n+1}}\, \mathrm{dt} = \frac{\lambda^{-(n-k+1)}}{\binom{n}{k}\, k} ,$$

we finally obtain for the expectation values of the sequential extinction times

$$E(\mathcal{T}_k) = \int_0^\infty \frac{\mathrm{d}H_k(t)}{\mathrm{dt}}\, t\, \mathrm{dt} = \frac{n-k}{k} \cdot \frac{1}{\lambda} ,\ n \geq k \geq 1 , \tag{10.10}$$

and $E(\mathcal{T}_0) = \infty$ (see Fig. 10.4). It is worth recognizing here another paradox of probability theory: Although extinction is certain, the expectation value for the time to extinction diverges. Similarly as the expectation values, we calculate the variances of the sequential extinction times:

$$\sigma^2(\mathcal{T}_k) = \frac{n(n-k)}{k^2(k-1)} \cdot \frac{1}{\lambda^2} ,\ n \geq k \geq 2 , \tag{10.11}$$

from which we see that the variances diverges for $k = 0$ and $k = 1$.

For distinct birth parameters, $\lambda_1, \ldots, \lambda_n$, and different initial numbers of individuals, $x_1(0), \ldots, x_n(0)$, the expressions for the expectation values become considerably more complicated, but the main conclusion remains unaffected: $E(\mathcal{T}_1)$ is finite whereas $E(\mathcal{T}_0)$ diverges.

10.1.2 Boundaries in one step birth-and-death processes

One step birth-and-death processes have been studied extensively and analytical solutions are available in table form [118]. For transition probabilities at most linear in n, $w^+(n) = \nu + \lambda n$ and $w^-(n) = \rho + \mu n$, one distinguishes birth (λ), death (μ), immigration (ν), and emigration (ρ) terms. Analytical solutions for the probability distributions were derived for all one step birth-and-death processes whose transitions probabilities are constant or maximally linear in the numbers of individuals n.

It is necessary, however, to consider also the influence of boundaries on these stochastic processes. For this goal we define an interval $[a, b]$ as domain of the stochastic variable $\mathcal{N}(t)$. Here we are dealing with classes of boundary conditions, *absorbing* and *reflecting boundaries*. In the former case, a particle that left the interval is not allowed to return, whereas the latter class of boundary implies that it is forbidden to exit from the interval. Boundary conditions can be easily implemented by *ad hoc* definitions of transition probabilities:

	Reflecting	Absorbing
Boundary at a	$w^-(a) = 0$	$w^+(a-1) = 0$
Boundary at b	$w^+(b) = 0$	$w^-(b+1) = 0$

The reversible chemical reaction with $w^-(n) = k_1\, n$ and $w^+(n) = k_2\,(n_0 - n)$, for example, had two reflecting barriers at $a = 0$ and $b = n_0$. Among the examples we have studied so far we were dealing with an absorbing boundary in the replication-extinction process between $\mathcal{N} = 1$ and $\mathcal{N} = 0$ that is tantamount to the lower barrier at $a = 1$ fulfilling $w^+(0) = 0$: The state $n = 0$ is the end point or ω-limit of all trajectories reaching it.

Compared, for example, to an unrestricted random walk on positive and negative integers, $n \in \mathbb{Z}$, a chemical reaction or a biological process has to be restricted by definition, $n \in \mathbb{N}^0$, since negative particle numbers are not allowed. In general, the one step birth-and-death master Equ. (9.7),

$$\frac{\partial P_n(t)}{\partial t} \;=\; w^+(n-1)\,P_{n-1}(t) + w^-(n+1)\,P_{n+1}(t) - \Big((w^+(n)+w^-(n)\Big)\,P_n(t)\;,$$

is not restricted to $n \in \mathbb{N}^0$ and thus does not automatically fulfil the proper boundary conditions to model a chemical reaction. A modification of the equation at $n = 0$ is required, which introduces a proper boundary of the process:

$$\frac{\partial P_0(t)}{\partial t} \;=\; w^-(1)\,P_1(t) \;-\; w^+(0)\,P_0(t)\;. \qquad (9.7')$$

This occurs *naturally* if $w^-(n)$ vanishes for $n = 0$, which is always the case when the constant term referring to migration vanishes, $\nu = 0$. With $w^-(0) = 0$ we only need to make sure that $P_{-1}(t) = 0$ and obtain Equ. (9.7'). This will be so whenever we take an initial state with $P_n(0) = 0\,\forall\, n < 0$, and it is certainly true for our conventional initial condition, $P_n(0) = \delta_{n,n_0}$ with $n_0 \geq 0$. By the same token we prove that the upper reflecting boundary for chemical reactions, $b = n_0$, fulfils the conditions of being *natural* too. Equipped with natural boundary conditions the stochastic process can be solved for the entire integer range, $n \in \mathbb{Z}$, and this is often much easier than with *artificial* boundaries. All the barriers we have encountered so far were natural.

An overview over a few selected birth-and-death processes is given in tables 10.1 and 10.2. Commonly, unrestricted and restricted processes are distinguished [118]. An unrestricted process is characterized by the possibility to reach all states $\mathcal{N}(t) = n$. A requirement imposed by physics demands that all changes in state space are finite for finite times, and hence the probabilities to reach infinity at finite times must vanish: $\lim_{n\to\pm\infty} P_{n,n_0} = 0$. The linear birth and death process in table 10.1 is unrestricted only in the positive direction and the state $\mathcal{N}(t) = 0$ is special because it represents an absorbing barrier. The restriction is here hidden and met by the condition $P_{n,n_0}(t) = 0\ \forall\ n < 0$.

Table 10.1: **Comparison of of results for some unrestricted processes.** Data are taken from [118, pp.10,11]. Abbreviation and notations: $\gamma \equiv \lambda/\mu$, $\sigma \equiv e^{(\lambda-\mu)t}$, $(n, n_0) \equiv \min\{n, n_0\}$, and $I_n(x)$ is the modified Bessel function.

Process	λ_n	μ_n	$g_{n_0}(s,t)$	$P_{n,n_0}(t)$	Mean	Variance	Ref.
Poisson	ν	0	$s^{n_0}\, e^{\nu(s-1)\,t}$	$\frac{(\nu t)^{n-n_0}\, e^{\nu\, t}}{(n-n_0)!}\,,\ n \geq n_0;\ n_0 > (0,n)$	$n_0 + \nu t$	νt	[39]
Poisson	0	ρ	$s^{n_0}\, e^{\rho(1-s)\,t/s}$	$\frac{(\rho t)^{n-n_0}\, e^{\rho\, t}}{(n_0-n)!}\,,\ n \leq n_0;\ n_0 < (0,n)$	$n_0 - \rho t$	ρt	[39]
	ν	ρ	$s^{n_0}\, e^{-(\nu+\rho)t+(\nu s+\rho/s)t}$	$\binom{\nu}{\rho}^{(n-n_0)/2} I_{n_0-n}(2t\sqrt{\nu\rho})\, e^{-(\nu+\rho)t}$	$n_0 + (\nu-\rho)t$	$(\nu+\rho)t$	[141]
Birth	λn	0	$\Big(1 - e^{\lambda t}(1-1/s)\Big)^{-n_0}$	$\binom{n}{n_0}\, e^{-n_0\lambda\, t}(1-e^{-\lambda\, t})^{n-n_0}\,,\ n \geq n_0;\ n_0 > (0,n)$	$n_0\, e^{\lambda t}$	$n_0\, e^{\lambda t}(e^{\lambda t}-1)$	[10]
Death	0	μn	$\Big(1 - e^{-\mu t}(1-s)\Big)^{n_0}$	$\binom{n_0}{n}\, e^{-n\mu\, t}(1-e^{-\mu\, t})^{n_0-n}\,,\ n \leq n_0;\ n_0 < (0,n)$	$n_0\, e^{-\mu t}$	$n_0\, e^{-\mu t}(1-e^{-\mu t})$	[10]
	ν	μn	$\Big(1 - e^{-\mu t}(1-s)\Big)^{n_0} \times$ $\times\, \exp\Big(\nu(s-1)(1-e^{-\mu t})/\mu\Big)$	$\exp\Big(-\frac{\nu}{\mu}(1-e^{-\mu t})\Big) \times$ $\times \sum_{k=0}^{(n,n_0)} \frac{e^{-\mu t k}(1-e^{-\mu t})^{n+n_0-2k}}{(n-k)!} \Big(\frac{\nu}{\mu}\Big)^{n-k}$	$n_0\, e^{-\mu t} +$ $+\frac{\nu(1-e^{-\mu t})}{\mu}$	$\Big(\frac{\nu}{\mu} + n_0 e^{-\mu t}\Big) \times$ $\times(1-e^{-\mu t})$	[39]
Birth& Death	λn	μn	$\Big(\frac{(\sigma-1)+(\gamma-\sigma)s}{(\gamma\sigma-1)\gamma(1-\sigma)s}\Big)^{n_0}$	$\gamma^n \sum_{k=0}^{(n,n0)} (-1)^k \binom{n+n_0-k-1}{n-k}\binom{n_0}{k} \times$ $\times \Big(\frac{1-\sigma}{1-\gamma\sigma}\Big)^{n+n_0-k} \Big(\frac{1-\sigma/\gamma}{1-\sigma}\Big)^{k}$	$n_0\, \sigma$	$\frac{n_0\sigma(\gamma+1)(\sigma-1)}{\gamma-1}$	[10]
	λn	λn	$\Big(\frac{\lambda t+(1-\lambda t)s}{1+\lambda t-\lambda t\, s}\Big)^{n}$	$\Big(\frac{\lambda t}{1+\lambda t}\Big)^{n+n_0} \sum_{k=0}^{(n,n_0)} \binom{n_0}{k} \times$ $\times \binom{n+n_0-k-1}{n-k} \Big(\frac{1-\lambda^2 t^2}{\lambda^2 t^2}\Big)^{k}$	n_0	$2n_0\lambda\, t$	

Table 10.2: **Comparison of of results for some restricted processes.** Data are taken from [118, pp.16,17]. Abbreviation and notations used in the table are: $\gamma \equiv \lambda/\mu$, $\sigma \equiv e^{(\lambda-\mu)t}$, $\alpha \equiv (\nu/\rho)^{(n-n_0)/2} e^{(\nu+\rho)t}$; $I_n = I_{-n} \equiv I_n\big(2(\nu\rho)^{1/2}t\big)$ where $I_n(x)$ is a modified Bessel function; $G_n \equiv G_n(\xi_j,\gamma)$ where G_n is a Gottlieb polynomial, $\hat{G}_n \equiv G_n(\hat{\xi}_j,\gamma)$, $G_n(x,\gamma) \equiv \gamma^n \sum_{k=0}^{n}(1-\gamma^{-1})^k \binom{n}{k}\binom{x-k+1}{k} = \gamma^n F(-n,-x,1,1-\gamma^{-1})$ where F is a hypergeometric function, ξ_j and $\hat{\xi}_j$ are the roots of $G_{u-l}(\xi_j,\gamma)=0$, $j=0,\ldots,u-l-1$ and $G_{u-l+1}(\hat{\xi}_j,\gamma) = \gamma\, G_{u-l}(\hat{\xi}_j,\gamma)$, $j=0,\ldots,u-l$, respectively; $H_n \equiv H_n(\zeta_j,\gamma)$, $\hat{H}_n \equiv H_n(\hat{\zeta}_j,\gamma)$, $H_n(x,\gamma) = G_n(x,\gamma^{-1})$, $H_{u-l}(\zeta_j,\gamma)=0$, $j=0,\ldots,u-l-1$ and $H_{u+l-1}(\hat{\zeta}_j,\gamma) = H_{u-l}(\hat{\zeta}_j,\gamma)/\gamma$, respectively.

λ_n	μ_n	Boundaries	$P_{n,n_0}(t)$	Ref.
ν	ρ	$u:$ abs; $l:-\infty$	$\alpha\big(I_{n-n_0} - I_{2u-n-n_0}\big)$	[39, 210]
ν	ρ	$u:+\infty$; $l:$ abs	$\alpha\big(I_{n-n_0} - I_{n+n_0-2l}\big)$	[39, 210]
ν	ρ	$u:$ refl; $l:-\infty$	$\alpha\left(I_{n-n_0} + \left(\frac{\nu}{\rho}^{1/2} I_{2u+l-n-n_0} + \left(1-\frac{\rho}{\nu}\right)\right)\cdot \sum_{j=2}^{\infty}\left(\frac{\nu}{\rho}\right)^{j/2} I_{2u-n-n_0+j}\right)$	[39, 210]
ν	ρ	$u:+\infty$; $l:$ refl	$\alpha\left(I_{n-n_0} + \left(\frac{\nu}{\rho}^{1/2} I_{n+n_0+l-2u} + \left(1-\frac{\rho}{\nu}\right)\right)\cdot \sum_{j=2}^{\infty}\left(\frac{\nu}{\rho}\right)^{j/2} I_{n+n_0-2l+j}\right)$	[39, 210]
ν	ρ	$u:$ abs; $l:$ abs	$\alpha\left(\sum_{k=-\infty}^{\infty} I_{n-n_0+2k(u-l)} - \sum_{k=0}^{\infty}\big(I_{n+n_0-2l+2k(u-l)} + I_{2l-n-n_0+2k(u-l)}\big)\right)$	[39, 210]
$\lambda(n-l+1)$	$\mu(n-l)$	$u:$ abs; $l:$ refl	$\gamma^{l-n}\sum_{k=0}^{u-l-1} G_{n_0-l}G_{n-l}\,\sigma^{\xi_k}\left(\sum_{j=0}^{u-l-1}\frac{G_j}{\gamma^j}\right)^{-1}$	[211, 262]
$\lambda(n-l+1)$	$\mu(n-l)$	$u:$ refl; $l:$ refl	$\gamma^{l-n}\sum_{k=0}^{u-l} \hat{G}_{n_0-l}\hat{G}_{n-l}\,\sigma^{\hat{\xi}_k}\left(\sum_{j=0}^{u-l}\frac{\hat{G}_j}{\gamma^j}\right)^{-1}$	[211, 262]
$\lambda(u-n)$	$\mu(u-n+1)$	$u:$ refl; $l:$ abs	$\gamma^{u-n}\sum_{k=0}^{u-l-1} H_{u-n_0}H_{u-n}\sigma^{-\zeta_k}\left(\sum_{j=0}^{u-l-1} H_j\gamma^j\right)^{-1}$	[211, 262]
$\lambda(u-n)$	$\mu(u-n+1)$	$u:$ refl; $l:$ refl	$\gamma^{u-n}\sum_{k=0}^{u-l} \hat{H}_{u-n_0}\hat{H}_{u-n}\sigma^{-\hat{\zeta}_k}\left(\sum_{j=0}^{u-l}\hat{H}_j\gamma^j\right)^{-1}$	[211, 262]

10.1.3 Branching processes in evolution

According to David Kendall's historical accounts on the centennial of the origins of stochastic thinking in population mathematics [171, 172] the name *branching process* was coined relatively late by Kolmogorov and Dmitriev in 1947 [180]. The interest in stochasticity of the evolution of reproducing populations, however, is much older. The origin of the problem is the genealogy of human males, which is reflected by the development of family names or *surnames* in the population. Commonly the stock of family names is *eroded* in the sense of steady disappearance of families in particular in small communities. The problem was clearly stated in a book by Alphonse de Candolle [47] and has been brought up by Sir Francis Galton after he had read de Candolle's book. The first rigorous mathematical analysis of a problem by means of a branching problem is commonly assigned to Galton and Reverend Henry William Watson [301] and the Galton-Watson process named after them has become a standard problem in branching processes. Apparently, Galton and Watson were not aware of earlier work on this topic [142], that had been performed almost thirty years earlier by Jules Bienaymé and reported in a publication [20]. Most remarkable Bienaymé discussed already the criticality theorem, which expresses different behavior of the Galton-Watson process for $m < 1$, $m = 1$, and $m < 1$, where m denotes the expected or mean number of sons per father. The three cases were called *subcritical*, *critical*, and *supercritical*, respectively, by Kolmogorov [179]. Watson's original work contained a serious error in the analysis of the supercritical case and this has been detected only by Johan Steffensen [275]. In the years after 1940 the Galton-Watson model received plenty of attention because of the analogies of genealogies and nuclear chain reactions. In addition, mathematicians became generally more interested in probability theory and stochasticity. The pioneering work related to nuclear chain reactions and criticality of nuclear reactors was done by Stan Ulam at the Los Alamos National Laboratory [82–85, 138]. Many other applications to biology and physics were found and branching processes have been studied intensively. By now, it seems, we have a clear picture on the Galton-Watson process and its history [172].

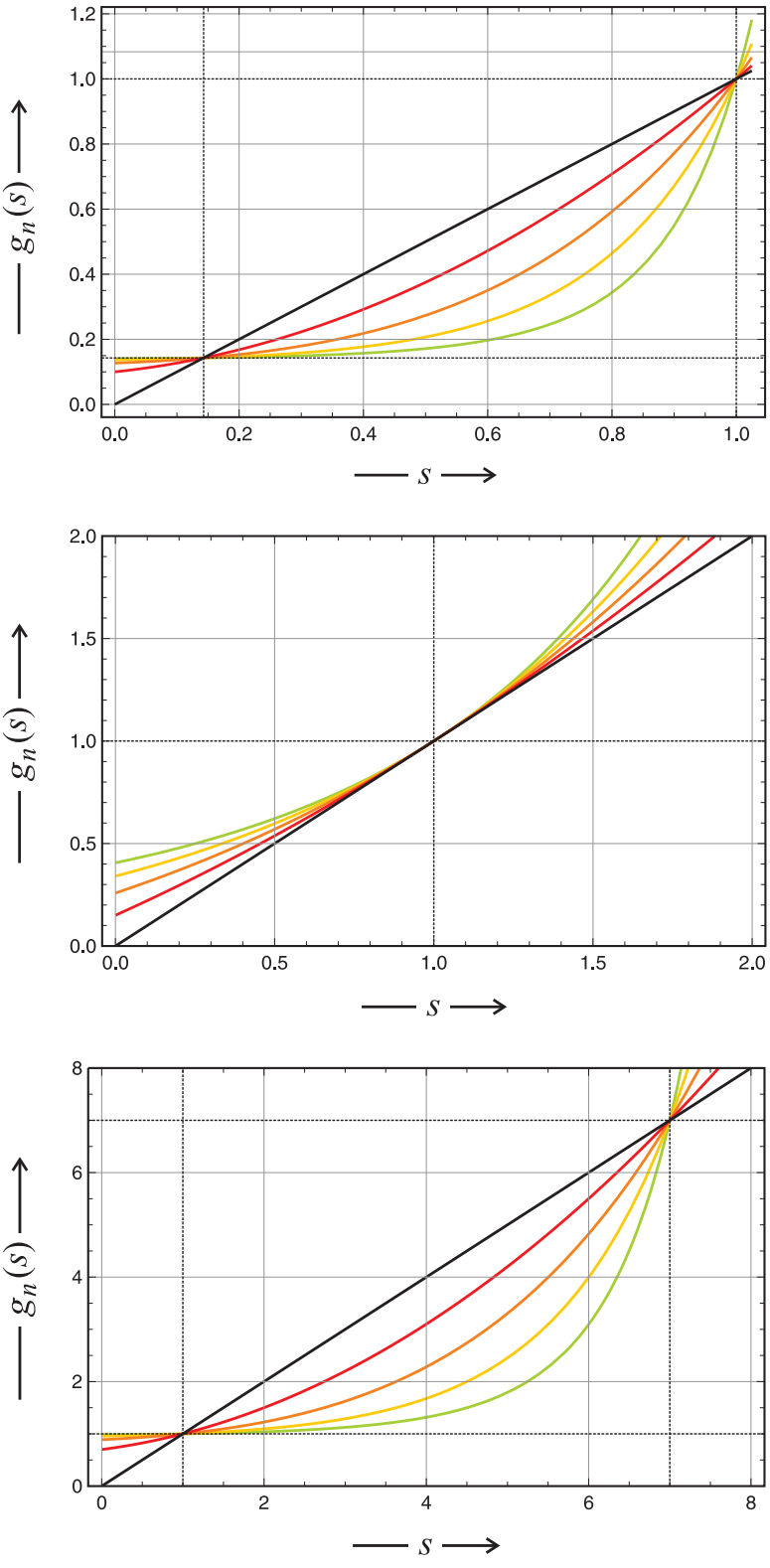


Figure 10.5: Continued on next page.

Figure 10.5: **Calculation of extinction probabilities for the Galton-Watson process.** The individual curves show the iterated generating functions of the Galton-Watson process, $g_0(s) = s$ (black), $g_1(s) = g(s) = p_0 + p_1 s + p_2 s^2$ (red), $g_2(s)$ (orange), $g_3(s)$ (yellow), and $g_4(s)$ (green), for different probability densities $\boldsymbol{p} = (p_0, p_1, p_2)$. Choice of parameters: supercritical case (upper part) $\boldsymbol{p} = (0.1, 0.2, 0.7)$, $m = 1.6$; critical case (middle part) $\boldsymbol{p} = (0.15, 0.7, 0.15)$, $m = 1$; subcritical case (lower part) $\boldsymbol{p}$=(0.7,0.2,0.1), $m = 0.4$.

The Galton-Watson process. A *Galton-Watson process* [301] deals with the generation of objects from objects of the same kind in the sense of reproduction. These objects can be neutrons, bacteria, or higher organisms, or men as in the family name genealogy problem. The Galton-Watson process is the simplest possible description of consecutive reproduction and falls into the class of branching processes. Recorded are only the population sizes of successive generations, which are considered as random variables: $\mathcal{Z}_0, \mathcal{Z}_1, \mathcal{Z}_2, \ldots$. A question of interest is the extinction of a population in generation n, and this simply means $\mathcal{Z}_n = 0$ from which follows that all the random variable is zero in all future generations: $\mathcal{Z}_{n+1} = 0$ if $\mathcal{Z}_n = 0$. Indeed, the extinction or disappearance of aristocratic family names was the problem that Galton wanted to model by means of a stochastic process. In the following presentation and analysis we make use of the two books [8, 135].

In mathematical terms the Galton-Watson process is a Markov chain $(\mathcal{Z}_n;\, n \in \mathbb{N}_0)$ on the nonnegative integers. The transition probabilities are defined in terms of a given probability function $\text{Prob}\,\{\mathcal{Z}_1 = k)\} = p_k;\, k \in \mathbb{N}_0$ with $p_k \leq 0, \sum p_k = 1$ have the

$$P(i,j) \;=\; \text{Prob}\,\{\mathcal{Z}_{n+1} = j | \mathcal{Z}_n = i\} \;=\; \begin{cases} p_j^{*i} & \text{if } \; i \geq 1,\, j \geq 0\,, \\ \delta_{0,j} & \text{if } \; i = 0,\, j \geq 0\,, \end{cases} \tag{10.12}$$

wherein δ_{ij} is the *Kronecker delta*[5] and $\{p_k^{*i};\, k \in \mathbb{N}_0\}$ is the i-fold convolution of $\{p_k;\, k \in \mathbb{N}_0\}$, and accordingly the probability mass function $f(k) = p_k$ is the only datum of the process. The use of the convolution of the probability distribution is an elegant mathematical trick for the rigorous analysis of the

[5]The Kronecker delta is named after the German mathematician Leopold Konecker

problem. Convolutions in explicit form are quite difficult to handle as we shall see in the case of the generating function. Nowadays one can use computer assisted symbolic computation but in Galton's times, in the 19th century handling of higher convolutions was quite hopeless.

The process describes an evolving population of particles or individuals and it might be useful although not necessary to define a time axis. The process starts with $\mathcal{Z}_0$ particles at time $T = 0$, each if which produces – independently of the others – a random number of offspring offspring at time $T = 1$ according to the probability density $f(k) = p_k$. The total number of particles in the first generation, $\mathcal{Z}_1$ is the sum of all $\mathcal{Z}_0$ random variables where each was drawn according to the pmf $f(p_k)$. The first generation produces $\mathcal{Z}_2$ particles at time $T = 2$, the second generation gives rise to the third with $\mathcal{Z}_3$ particles at time $T = 3$, and so on. Since discrete times T_n are equivalent to the numbers of generations n we shall refer only to generations ion the following. From (10.12) follows that the future development of the process at any time is independent of the history and this constitutes the Markov property.

The number of offspring produced by a single parent particle in the n-th generation is a random variable $\mathcal{Z}_n^{(1)}$ where the superscript indicates $\mathcal{Z}_0 = 1$. In general we shall write for the branching process $(\mathcal{Z}_n^{(i)}\,;\ n \in \mathbb{N}_0)$ when we want to express that the process started with i particles. Since $i = 1$ is the by far most common case, we write simply $\mathcal{Z}_n^{(1)} = \mathcal{Z}_n$. Equ. (10.12) tells that $\mathcal{Z}_{n+k} = 0\,\forall\, k \geq 0$. Accordingly, the state $\mathcal{Z} = 0$ is absorbing and reaching $\mathcal{Z} = 0$ is tantamount to becoming extinct.

In order to analyze the process we shall make use of the probability generating function

$$g(s) \;=\; \sum_{k=0}^{\infty} p_k\, s^k\,,\ |s| \leq 1\ , \tag{10.13}$$

---

and represents the discrete analogue of Dirac's delta function:

$$\delta_{ij} \;=\; \begin{cases} 0 & \text{if}\ \ i \neq j\,, \\ 1 & \text{if}\ \ i = j\,. \end{cases}$$

where s is complex in general but we shall assume here $s \in \mathbb{R}^1$. In addition, we define the iterates of the generating function:

$$g_0(s) \;=\; s\,,\; g_1(s) \;=\; g(s)\,,\; g_{n+1}(s) \;=\; g\big(g_n(s)\big)\,,\; n = 1, 2, \ldots\;. \qquad (10.14)$$

Expressed in terms of transition probabilities the generating function is of the form

$$\sum_{j=0}^{\infty} P(1,j)\, s^j \;=\; g(s) \;\text{ and }\; \sum_{j=0}^{\infty} P(i,j)\, s^j \;=\; \big(g(s)\big)^i\,,\; i \geq 1\;. \qquad (10.15)$$

Denoting the n-step transition probability by $P_n(i,j)$ and using the Chapman-Kolmogorov equation we obtain

$$\begin{aligned}
\sum_{j=0}^{\infty} P_{n+1}(1,j)\, s^j \;&=\; \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} P_n(1,k)\, P(k,j)\, s^j \;= \\
&=\; \sum_{k=0}^{\infty} P_n(1,k) \sum_{j=0}^{\infty} P(k,j)\, s^j \;= \\
&=\; \sum_{k=0}^{\infty} P_n(1,k)\, \big(g(s)\big)^k\;.
\end{aligned}$$

Writing $g_{(n)} = \sum_j P_n(1,j)\, s^j$ the last equation has shown that

$$g_{(n+1)}(s) \;=\; g_{(n)}\big(g(s)\big)$$

which yields the fundamental relation

$$g_{(n)}(s) \;=\; g_n(s)\;, \qquad (10.16)$$

and by making use of Equ. (10.15) we find

$$\sum_{j=0}^{\infty} P_n(i,j)\, s^j \;=\; \Big(g_n(s)\Big)^i\;. \qquad (10.17)$$

Equ. (10.16) expressed as "The generating function of $\mathcal{Z}_n$ is the n-iterate $g_n(s)$", provides a tool for the calculation of the generating function. As stated in Equ. (10.12) the probability distribution of $\mathcal{Z}_n$ is obtained as the

n-th convolution or iterate of $g(s)$. The explicit form of an n-th convolution is hard to compute and the true value of (10.16) lies in the calculation of the moments of $\mathcal{Z}_n$ and in the possibility to derive asymptotic laws for large n.

For the purpose of illustration we present the first iterates of the simplest useful generating function

$$g(s) \;=\; p_0 \,+\, p_1\, s \,+\, p_2\, s^2 \;.$$

The first convolution $g_2(s) = g\big(g(s)\big)$ contains ten terms already:

$$\begin{aligned} g_2(s) \;=\;& p_0 + p_0p_1 + p_0^2p_2 \,+\, (p_1^2 + 2p_0p_1p_2)\, s \,+ \\ & +\, (p_1p_2 + p_1^2p_2 + 2p_0p_2^2)\, s^2 \,+\, 2p_1p_2^2\, s^3 \,+\, p_2^3\, s^4. \end{aligned}$$

The next convolution, $g_3(s)$, contains already nine constant terms that contribute to the probability of extinction $g_n(0)$, and $g_4(s)$ already 29 terms.

It is straightforward to compute the moments of the probability distributions from the generating function:

$$\frac{\partial g(s)}{\partial s} \;=\; \sum_{k=0}^{\infty} k\, p_k\, s^{k-1} \;\text{ and }\; \frac{\partial g(s)}{\partial s}\,\Big|_{s=1} = \; E(\mathcal{Z}_1) \;=\; m \;, \tag{10.18a}$$

$$\frac{\partial^2 g(s)}{\partial s^2} \;=\; \sum_{k=0}^{\infty} k(k-1)\, p_k\, s^{k-2} \;\text{ and }\; \frac{\partial^2 g(s)}{\partial s^2}\,\Big|_{s=1} = \; E(\mathcal{Z}_1^2) \,-\, m \;,$$

$$\operatorname{var}(\mathcal{Z}_1) \;=\; \frac{\partial^2 g(s)}{\partial s^2}\,\Big|_{s=1} + m \,-\, m^2 \;=\; \sigma^2 \;. \tag{10.18b}$$

Next we calculate the moments of the distribution in higher generations and differentiate the last expression in Equ. 10.14 at $|s| = 1$:

$$\begin{aligned} \frac{\partial g_{n+1}(s)}{\partial s}\,\Big|_{s=1} \;&=\; \frac{\partial g(s)}{\partial s}\Big(g_n(s)\,\Big|_{s=1}\Big)\,\frac{\partial g_n(s)}{\partial s}\,\Big|_{s=1} = \\ &=\; \frac{\partial g(s)}{\partial s}\,\Big|_{s=1}\,\frac{\partial g_n(s)}{\partial s}\,\Big|_{s=1} \;\text{ and} \\ E(\mathcal{Z}_{n+1}) \;&=\; E(\mathcal{Z})\,E(\mathcal{Z}_n) \;\text{ or }\; E(\mathcal{Z}_n) \;=\; m^n \;, \end{aligned} \tag{10.19}$$

by induction. Provided the second derivative of the generating function at $|s| = 1$ is finite, Equ. 10.14 can be differentiated twice:

$$\frac{\partial^2 g_{n+1}(s)}{\partial s^2}\,\Big|_{s=1} = \; \frac{\partial g(s)}{\partial s}\,\Big|_{s=1}\,\frac{\partial^2 g_n(s)}{\partial s^2}\,\Big|_{s=1} + \frac{\partial^2 g(s)}{\partial s^2}\,\Big|_{s=1}\Big(\frac{\partial g_n(s)}{\partial s}\,\Big|_{s=1}\Big)^2 \;,$$

and $\partial^2 g(s) \,/\, \partial s^2 \,\big|_{s=1}$ is obtained by repeated application. The final result is:

$$\mathrm{var}(\mathcal{Z}_n) \;=\; E(\mathcal{Z}_n^2) \,-\, E(\mathcal{Z}_n)^2 \;=\; \begin{cases} \frac{\sigma^2\, m^n\, (m^n-1)}{m\,(m-1)} & ,\ \text{if } m \neq 1 \\ n\,\sigma^2 & ,\ \text{if } m = 1 \end{cases} . \qquad (10.20)$$

Thus, we have $E(\mathcal{Z}_n) = m^n$ and provided $\sigma = \mathrm{var}(\mathcal{Z}_1) < \infty$ the variances are given by Equ. (10.20).

Two more assumptions are made in order to facilitate the analysis: (i) Neither the probabilities p_0 and p_1 nor their sum are equal to one, $p_0 < 1$, $p_1 < 1$, and $p_0 + p_1 < 1$, and this implies that $g(s)$ is strictly convex on the unit interval $0 \leq s \leq 1$, and (ii) the expectation value $E(\mathcal{Z}_1) = \sum_{k=0}^{\infty} k\, p_k$ is finite, and from the finiteness of the expectation value follows $\partial g/\partial s|_{s=1}$ is finite too since $|s| \leq 1$.

Eventually we can now consider Galton's extinction problem of family names. The straightforward definition of extinction is given in terms of a random sequence $(\mathcal{Z}_n;\, n = 0, 1, 2, \ldots, \infty)$, which consists of zeros except a finite number of positive integer value at the beginning of the series. The random variable $\mathcal{Z}_n$ is integer valued and hence extinction is tantamount to the event $\mathcal{Z}_n \to 0$. From $P(\mathcal{Z}_{n+1} = 0 | \mathcal{Z}_n = 0) = 1$ follows the equality

$$\begin{aligned} P(\mathcal{Z}_n \to 0) \;&=\; P(\mathcal{Z}_n = 0 \;\text{ for some } n) \;= \\ &=\; P\big((\mathcal{Z}_1 = 0) \cup (\mathcal{Z}_2 = 0) \cup \cdots \cup (\mathcal{Z}_n = 0)\big) \;= \\ &=\; \lim_{n\to\infty} P\big((\mathcal{Z}_1 = 0) \cup (\mathcal{Z}_2 = 0) \cup \cdots \cup (\mathcal{Z}_n = 0)\big) \;= \\ &=\; \lim P(\mathcal{Z}_n = 0) \;=\; \lim g_n(0) \;, \end{aligned} \qquad (10.21)$$

and the fact that $g_n(0)$ is a nondecreasing function of n (see also Fig. 10.6).

We define a probability of extinction, $q = P(\mathcal{Z}_n \to 0) = \lim g_n(0)$ and show that $m = E(\mathcal{Z}_1) \leq 1$ the probability of extinction fulfils $q = 1$, and the family names disappear in finite time. For $m > 1$, however, the extinction probability is the unique solution less than one of the equation

$$s = g(s) \;\text{ for }\; 0 \leq s < 1 \;. \qquad (10.22)$$

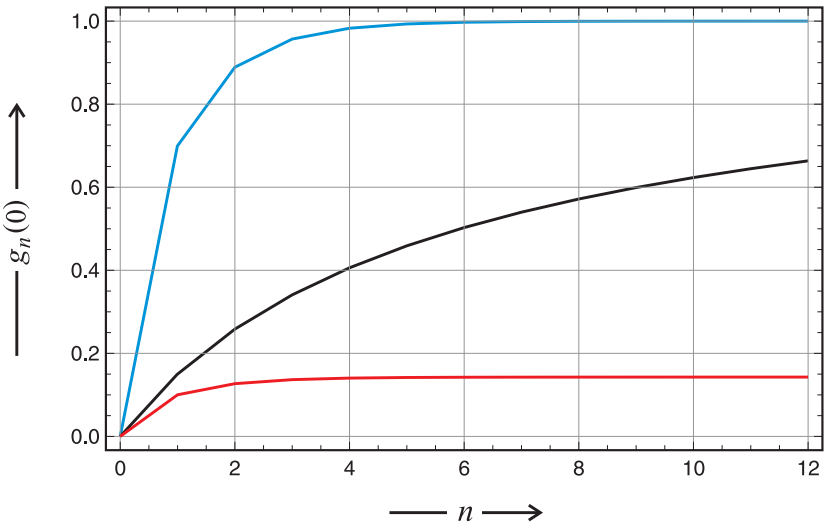


Figure 10.6: **Extinction probabilities in the Galton-Watson process.** Shown are the extinction probabilities for the three Galton-Watson processes discussed in Fig. 10.5. The supercritical process ($\boldsymbol{p} = (0.1, 0.2, 0.7)$, $m = 1.6$; red) is characterized by a probability of extinction of $q = \lim g_n < 1$ leaving room for a certain probability of survival, whereas both, the critical ($\boldsymbol{p} = (0.15, 0.7, 0.15)$, $m = 1$; black) and the subcritical process ($\boldsymbol{p} = (0.7, 0.2, 0.1)$, $m = 0.4$; blue) lead to certain extinction, $q = \lim g_n = 1$. In the critical case we observe much slower convergence than in the super- or subcritical case representing a nice example of *critical slowing down*.

It is straightforward to show by induction that $g_n(0) < 1$, $n = 0, 1, \ldots$. From Equ. (10.21) we know

$$0 = g_n(0) \leq g_1(0) \leq g_2(0) \leq \cdots \leq q = \lim g_n(0) \; .$$

Making use of the relations $g_{n+1}(0) = g\big(g_n(0)\big)$ and $\lim g_n(0) = \lim g_{n+1}(0) = q$ we derive $q = g(q)$ for $0 \leq q \leq 1$ – trivially fulfilled for $q = 1$ since $g(1) = 1$:

(i) $m \leq 1$, then $(\partial g(s)/\partial s) < 1$ for $0 \leq s < 1$. Next we use the *law of the mean*[6] express $g(s)$ in terms of $g(1)$ and for $m \leq 1$ we find $g(s) > s$ in the entire range $0 \leq s < 1$. There is only the trivial solution $q = g(q)$ with $q = 1$ and extinction is certain. □

[6]The law of the mean expresses the difference in the values of a function $f(x)$ in terms

(ii) $m > 1$, then $g(s) < s$ for s sightly less than one because $(\partial g/\partial s)|_{s=1} = m > 1$, whereas for $s = 0$ we have $g(0) > 0$ and hence we have at least one solution $s = g(s)$ in the half-open interval $[0, 1[$. Assume there were two solutions, for example s_1 and s_2 with $0 \leq s_1 < s_2 < 1$ than Rolle's theorem named after the French mathematician Michel Rolle would demand the existence of ξ and η with $s_1 < \xi < s_2 < \eta < 1$ such that $(\partial g(s)/\partial s)|_{s=\xi} = (\partial g(s)/\partial s)|_{s=\eta} = 1$ but this contradicts the fact that that $g(s)$ is strictly convex. In addition $\lim g_n(0)$ cannot be one because $(g_n(0); n = 0, 1, \ldots)$ is a nondecreasing sequence. If $g_n(0)$ were slightly less than one then $g_{n+1}(0) = g\big(g_n(0)\big)$ would be less than $g_n(0)$ and the series were decreasing. Accordingly, $q < 1$ is the unique solution of Equ. 10.22 in $[0, 1[$. □

The answer is simple and straightforward: When a father has on the average one son or less, the family name is doomed to disappear, when he has more than one son there is a finite probability of survival $0 < (1-q) < 1$, which, of course, increases with increasing expectation value m, the average number of sons. Reverend Henry William Watson correctly deduced that the extinction probability is given by a root of Equ. (10.22). He failed, however, to recognize that for $m > 1$ the relevant root is the one with $q < 1$ [105, 301]. It is remarkable that it took almost fifty years for the mathematical community to detect the error that has a drastic consequence for the result.

Replication and mutation as multitype branching process.

---

of the derivative at one particular point $x = x_1$ and the difference in the arguments

$$f(b) - f(a) = (b-a)\,(\partial f/\partial x)|_{x=x_1}\,,\ a < x < b\,.$$

The law of the mean is fulfilled at least at one point x_1 on the arc between a and b.

10.1.4 The Wright-Fisher and the Moran process

Here we shall introduce two common stochastic models in population biology, the *Wright-Fisher model* named after Sewall Wright and Ronald Fisher and the *Moran model* named after the Australian statistician Pat Moran. The Wright-Fisher model and the Moran model are stochastic models for evolution of allele distributions in populations with constant population size [22]. The first model [92, 318] also addressed as *beanbag* population genetics is presumably the simplest process for the illustration of genetic drift and definitely the most popular one [43, 88, 136, 198] deals with strictly separated generations, whereas the Moran process [212, 213] based on continuous time and overlapping generations is generally more appealing to statistical physicists. Both processes are introduced here for the simplest scenarios: haploid organisms, two alleles of the gene under consideration and no mutation. Extension to more complicated cases is readily possible. The primary question that was thought to be addressed by the two models is the evolution of populations in case of neutrality for selection.

The Wright-Fisher process. The Wright-Fisher process is illustrated in Fig. 10.7. A single reproduction event is modeled by a sequence of four steps: (i) A gene is randomly chosen from the gene pool of generation T containing exactly N genes distributed over m alleles, (ii) it is replicated, (iii) the original is put back into the gene pool T, and (iv) the copy is put into the gene pool of the next generation $T+1$. The process is terminated when the next generation gene pool has exactly N genes. Since filling the gene pool of the $T+1$ generation depends exclusively on the distribution of genes in the pool of generation T, and earlier gene distributions have no influence on the process the Wright-Fisher model is Markovian.

In order to simplify the analysis we assume two alleles A and B, which are present in a_T and b_T copies in the gene pool at generation T. Since the total number of genes is constant, $a_T + b_T = N$ and $b_T = N - a_T$, we are dealing with a single discrete variable, a_T, $T \in \mathbb{N}$. A new generation $T+1$ is produced from the gene pool at generation T through picking with

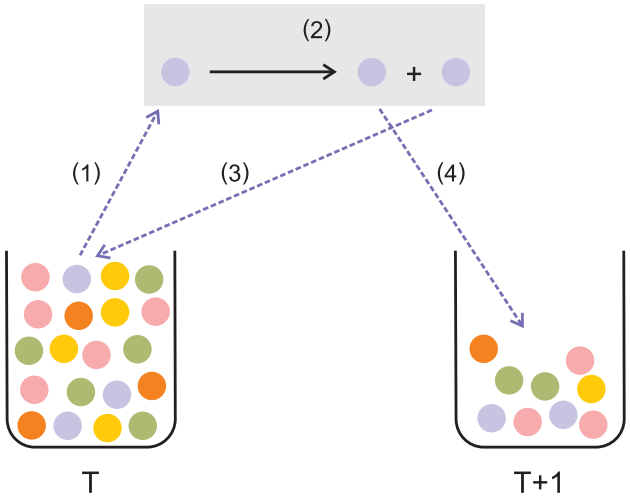


Figure 10.7: **The Wright-Fisher model of *beanbag* genetics.** The gene pool of generation T contains N gene copies chosen from m alleles. Generation $T+1$ is built from generation T through ordered cyclic repetition of a four step event: (1) random selection of one gene from the gene pool T, (2) error-free copying of the gene, (3) putting back the original into gene pool T, and (4) placing the copy into the gene pool of the next generation $T+1$. The procedure is repeated until the gene pool $T+1$ contains exactly N genes. No mixing of generations is allowed.

replacement N times a gene. The probability to obtain $n = a_{T+1}$ alleles A in the new gene pool is given by the binomial distribution:

$$\text{Prob}\,(a_{T+1} = n) \;=\; \binom{N}{n} p_{\mathsf{A}}^{n}\, p_{\mathsf{B}}^{N-n}\;,$$

$p_{\mathsf{A}} = a_T/N$ and $p_{\mathsf{B}} = b_T/N = (N-a_T)/N$ with $p_{\mathsf{A}}+p_{\mathsf{B}} = 1$ are the individual probabilities of picking A or B, respectively. The transition probability from m alleles A at time T to n alleles at time $T+1$ is simply given by[7,8]

$$p_{nm} \;=\; \binom{N}{n} \left(\frac{m}{N}\right)^{n} \left(1-\frac{m}{N}\right)^{N-n} \;. \tag{10.23}$$

[7]The notation applied here is the conventional way of writing transitions in physics: p_{nm} is the probability of the transition $n \leftarrow m$, whereas many mathematicians would write p_{mn} indicating $m \rightarrow n$.

[8]For doing actual calculations one has to recall the convention $0^0 = 1$ used in probability theory and combinatorics but commonly not in analysis where 0^0 is an indefinite expression.

Since the construction of th e gene pool at generation $T+1$ is fully determined by the gene distribution at generation T, the process is Markovian.

In order to study the evolution of a population an initial state has to be specified. We assume that the number of alleles A has been n_0 at generation $T = 0$ and accordingly we are calculating the probability $P(n, T|\, n_0, 0)$. Since the Wright-Fisher model does not contain any interactions between alleles or mutual dependencies between processes involving alleles, the process can be modeled by means of linear algebra. We define a probability vector $\boldsymbol{p}$ and a transition matrix, P:

$$\boldsymbol{P}(T) \;=\; \boldsymbol{p}(T) \;=\; \begin{pmatrix} p_0(T) \\ p_1(T) \\ p_2(T) \\ \vdots \end{pmatrix} \qquad \text{and} \qquad \mathrm{P} \;=\; \begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \;.$$

Conservation of probability provides two conditions: (i) The probability vector has be normalized, $\sum_n p_n(T) = 1$ and (ii) has to remain normalized in future generations, $\sum_n p_{nm} = 1$.[9] The evolution is now simply described by the matrix equation

$$\boldsymbol{p}(T+1) \;=\; \mathrm{P} \cdot \boldsymbol{p}(T) \quad \text{or} \quad \boldsymbol{p}(T) \;=\; \mathrm{P}^T \cdot \boldsymbol{p}(0) \;. \tag{10.24}$$

Equ. (10.24) is formally identical with the matrix formulation of linear *difference equations*, which are extensively discussed in [44, pp.179-216]. The solutions of (10.24) are commonly analyzed in terms of the eigenvalues of matrix P [89]

$$\lambda_k \;=\; \binom{N}{k} \frac{k!}{N^k}\,;\; k = 0, 1, 2, \ldots \;, \tag{10.25}$$

and the corresponding eigenvectors. In case of the long-time behavior we only consider the largest eigenvector or in case of degeneracy all eigenvectors belonging to the largest eigenvalue. Stationarity implies $\boldsymbol{p}(T+1) = \boldsymbol{p}(T) = \bar{\boldsymbol{p}}$, or $\mathrm{P}\,\bar{\boldsymbol{p}} = \bar{\boldsymbol{p}}$, a stationary probability distribution is an eigenvector $\bar{\boldsymbol{p}}$ associated with an eigenvalue $\lambda = 1$ of the transition matrix P.

[9] A matrix P with this property is called a stochastic matrix.

Here we are dealing with a case of degeneracy since the largest eigenvalue $\lambda = 1$ is obtained twice from Equ. (10.25): $\lambda_0 = \lambda_1 = 1$. The corresponding two eigenvectors are of the form

$$\begin{aligned} \boldsymbol{\zeta}_0 &= (1,0,0,0,0,\ldots,0)^{\mathrm{t}} \quad \text{and} \\ \boldsymbol{\zeta}_1 &= (0,0,0,0,0,\ldots,1)^{\mathrm{t}} \; . \end{aligned} \tag{10.26}$$

We have to recall that in case of degenerate eigenvalue any properly normalized linear combination of the eigenvectors is also a legitimate solution of the eigenvalue problem. Here we have to apply the L^1-norm and obtain

$$\boldsymbol{\eta} \;=\; \alpha\,\boldsymbol{\zeta}_0 \,+\, \beta\,\boldsymbol{\zeta}_1 \;\; \text{and} \;\; \alpha + \beta = 1 \; ,$$

and find for the general solution of the stationary state

$$\boldsymbol{\eta} \;=\; (1-\pi,0,0,0,0,\ldots,\pi)^{\mathrm{t}} \; . \tag{10.27}$$

The interpretation of the result is straightforward: The allele A becomes fixated in the population with probability π and is lost with probability $1-\pi$, and the Wright-Fisher model provides a simple explanation for gene fixation by random drift. What remains to be calculated is the value of π.[10] For this goal we make use of expectation vale of the number of alleles A and show first that it does not change with generations:

$$E(a_{T+1}) \;=\; \sum_n n p_n(T+1) \;=\; \sum_n n \sum_m p_{nm} p_m(T) \;=\; \sum_m m p_m(T) \;=\; E(a_T) \, ,$$

where we made use of the relation

$$\begin{aligned} \sum_{n=0}^{N} n\, p_{nm} \;&=\; \sum_{n=0}^{N} n \binom{N}{n} \left(\frac{m}{N}\right)^n \left(1-\frac{m}{N}\right)^{N-n} \\ &=\; \left(1-\frac{m}{N}\right)^N \sum_{n=0}^{N} n \binom{N}{n} \left(\frac{m}{N-m}\right)^n \;=\; m \; , \end{aligned}$$

[10] Although a stationary state does not depend on initial conditions in the nondegenerate case, this is not true for the linear combination of degenerate eigenvectors: α and β, and π are functions of the initial state.

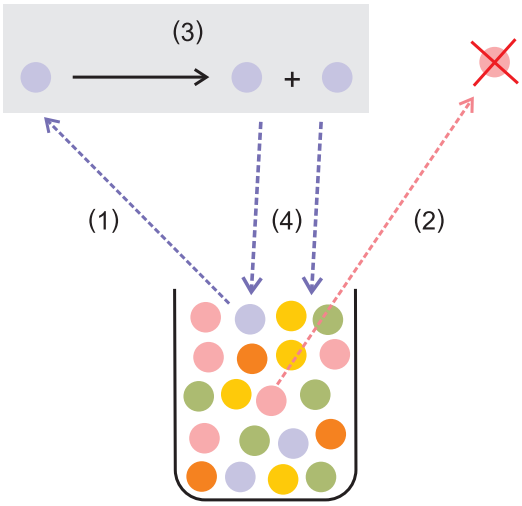


Figure 10.8: **The Moran process.** The Moran process is a continuous time model for the same problem handled by the Wright-Fisher model (Fig. 10.7). The gene pool of a population of N genes chosen from m alleles is represented by the urn in the figure. Evolution proceeds via successive repetition of a four step process: (1) One gene is chosen from the gene pool at random, (2) a second gene is randomly chosen and deleted, (3) the first gene is copied, and (4) both genes, original and copy, are put back into the urn. The Moran process has overlapping generations and moreover the notion of generation is not well defined.

which is solved readily by making use of the finite series

$$\sum_{n=0}^{N} \binom{N}{n} n\,\gamma^n \;=\; \gamma\,N\,(1+\gamma)^{N-1}\;.$$

From the generation independent expectation value we obtain π:

$$n_0 \;=\; E(a_T) \;=\; \lim_{T\to\infty} E(a_T) \;=\; N\,\pi\;,$$

and the probability for the fixation of A finally is n_0/N.

The Moran process. The Moran process introduced by Pat Moran [212] is a continuous time process and deals with transitions that are defined for single events. As in the Wright-Fisher model we are dealing with two alleles, A and B, and the probabilities for choosing A or B are p_{A} and p_{B}, respectively. Unlike

the Wright-Fisher model there is no defined previous generation from which a next generation is formed by sampling N genes. Overlapping generations make it difficult – if not impossible – to define generations unambiguously. The *event* in the Moran process is a *combined birth-and-death step*: Two genes are picked, one is copied and both template and copy are put back into the urn, and the second one is deleted (see Fig. 10.8). The probabilities are calculated form the state of the urn just before the event $p_{\mathsf{A}} = n(t)/N$ and $p_{\mathsf{B}} = \big(N - n(t)\big)/N$ where $n(t)$ is the number of alleles A, $N - n(t)$ the number of alleles B, and N is the constant total number of genes.

The transition matrix P of the Moran model is tridiagonal since only the changes $\Delta n = 0, \pm 1$ can occur: The number of alleles A, n, remains unchanged if two A's or two B's are chosen, if a pair $\mathsf{A} + \mathsf{B}$ is chosen and A dies n decreases by one, and eventually if A reproduces n goes up by one:

$$p_{nm} \;=\; \begin{cases} \left(1 - \frac{m}{N}\right) p_{\mathsf{A}}(m) = \left(1 - \frac{m}{N}\right)\left(\frac{m}{N}\right) , & \text{if} \quad n = m+1 \\ \left(\frac{m}{N}\right) p_{\mathsf{A}}(m) + \left(1 - \frac{m}{N}\right) p_{\mathsf{B}}(m) = \left(\frac{m}{N}\right)^2 + \left(1 - \frac{m}{N}\right)^2 , & \text{if} \quad n = m \\ \left(\frac{m}{N}\right) p_{\mathsf{B}}(m) = \left(\frac{m}{N}\right)\left(1 - \frac{m}{N}\right) , & \text{if} \quad n = m-1 \\ 0 , & \text{otherwise} , \end{cases}$$

and probability conservation, $\sum_n p_{nm} = 1$, is easily verified.

The transition matrix $\mathrm{P} = \{p_{nm}\}$ has tridiagonal form and eigenvalues and eigenvectors are readily calculated [212, 213]:

$$\lambda_k \;=\; 1 \,-\, \frac{k(k-1)}{N^2} ; \; k = 0, 1, 2, \ldots \; , \tag{10.28}$$

The first two eigenvectors are the same as found for the Wright-Fisher-model. The third eigenvector can be used to calculate the evolution towards fixation:

$$\boldsymbol{p}(t) \;\approx\; \begin{pmatrix} 1 - \frac{n_0}{N} \\ 0 \\ \vdots \\ 0 \\ \frac{n_0}{N} \end{pmatrix} + \frac{6 n_0 (N - n_0)}{N(N^2 - 1)} \begin{pmatrix} -\frac{N-1}{2} \\ 1 \\ \vdots \\ 1 \\ -\frac{N-1}{1} \end{pmatrix} \left(1 - \frac{2}{N^2}\right)^t \; ,$$

where n_0 is the initial number of A alleles [22]. The stationary state is identical with the fixation equilibrium in the Wright-Fisher model.

10.2 Master equations in biology

10.3 Neutrality and Kimura's theory of evolution

11. Perspectives and Outlook

Bibliography

[1] M. Abramowitz and I. A. Segun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications, New York, 1965.

[2] R. Aguirre-Hernandéz, H. H. Hoos, and A. Condon. Computational RNA secondary structure design: Empirical complexity and improved methods. *BMC Bioinformatics*, 8:e34, 2007.

[3] T. Aita, Y. Hayashi, H. Toyota, Y. Husimi, I. Urabe, and T. Yomo. Extracting characteristic properties of fitness landscape from *in vitro* molecular evolution: A case study on infectivity of *fd* phage to *e. coli*. *J. Theor. Biol.*, 246:538–550, 2007.

[4] E. Akin. *The Geometry of Population Genetics*, volume 31 of *Lecture Notes in Biomathematics*. Springer-Verlag, Berlin, 1979.

[5] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, Taylor & Francis Group, New York, fifth edition, 2008.

[6] L. W. Ancel and W. Fontana. Plasticity, evolvability, and modularity in RNA. *J. Exp. Zool.(Mol. Dev. Evol.)*, 288:242–283, 2000.

[7] L. W. Ancel-Meyers and W. Fontana. Evolutionary lock-in and the origin of modularity in RNA structure. In W. Callebaut and D. Rasskin-Gutman, editors, *Modularity – Understanding the Development and Evolution of Natural Complex Systems*, pages 129–141. MIT Press, Cambridge, MA, 2005.

[8] K. B. Athreya and P. E. Ney. *Branching Processes*. Springer-Verlag, Heidelberg, DE, 1972.

[9] J. F. Atkins, R. F. Gesteland, and T. R. Cech, editors. *RNA Worlds. From Life's Origins to Diversity in Gene Regulation*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2011.

[10] N. T. J. Bailey. *The Elements of Stochastic Processes with Application in the Natural Sciences*. Wiley, New York, 1964.

[11] T. S. Bayer and C. D. Smolke. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat. Biotechnol.*, 23:337–343, 2005.

[12] G. W. Beadle and E. L. Tatum. Genetic control of biochemical reactions in *neuspora*. *Proc. Natl. Acad. Sci. USA*, 27:499–506, 1941.

[13] A. Becskei, B. Seraphin, and L. Serrano. Positive feedback in eukaryotic gene networks: Cell differentiation by graded to binary response conversion. *EMBO J.*, 20:2528–2535, 2001.

[14] C. K. Biebricher. Darwinian selection of self-replicating RNA molecules. In M. K. Hecht, B. Wallace, and G. T. Prance, editors, *Evolutionary Biology, Vol. 16*, pages 1–52. Plenum Publishing Corporation, 1983.

[15] C. K. Biebricher. Quantitative analysis of mutation and selection in self-replicating RNA. *Adv. Space Res.*, 12(4):191–197, 1992.

[16] C. K. Biebricher and M. Eigen. The error threshold. *Virus Research*, 107:117–127, 2005.

[17] C. K. Biebricher, M. Eigen, and J. William C. Gardiner. Kinetics of RNA replication. *Biochemistry*, 22:2544–2559, 1983.

[18] C. K. Biebricher, M. Eigen, and J. William C. Gardiner. Kinetics of RNA replication: Plus-minus asymmetry and double-strand formation. *Biochemistry*, 23:3186–3194, 1984.

[19] C. K. Biebricher, M. Eigen, and J. William C. Gardiner. Kinetics of RNA replication: Competition and selection among self-replicating RNA species. *Biochemistry*, 24:6550–6560, 1985.

[20] I. J. Bienaymé. "da la loi de multiplication et de la durée des familles". *Soc. Philomath. Paris Extraits*, Ser. 5:37–39, 1845.

[21] J. Binet. "mémoire sur l'intégration des équations linéaires aux différences finies, d'un ordre quelconque, à coefficients variables". *Comptes Rendue Hebdomadaires des Séances de l'Académie des Sciences (Paris)*, 17:559–567, 1843.

[22] R. A. Blythe and A. J. McKane. Stochastic models of evolution in genetics, ecology and linguistics. *J. Stat. Mech,: Theor. Exp.*, 2007. P07018.

[23] M. N. Boddy, P. L. Gaillard, W. H. McDonald, P. Shanahan, J. R. Yates 3rd, and P. Russell. Mus81-Eme1 are essential components of a holliday junction resolvase. *Cell*, 107:537–548, 2001.

[24] P. M. Boyle and P. A. Silver. Harnessing nature's toolbox: Regulatory elements for synthetic biology. *J. Roy. Soc. Interface*, 6:S535–S546, 2009.

[25] S. Brakmann and K. Johnsson, editors. *Directed Molecular Evolution of Proteins or How to Improve Enzymes for Biocatalysis*. Wiley-VCH, Weinheim, DE, 2002.

[26] S. Brenner. Theoretical biology in the third millenium. *PhilṪrans. Roy. Soc. London B*, 354:1963–1965, 1999.

[27] S. Brenner. Hunters and gatherers. *The Scientist*, 16(4):14, 2002.

[28] R. Brown. 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.

[29] V. Bryson and W. Szybalski. Microbial selection. *Science*, 116:45–46, 1952.

[30] J. J. Bull, L. Ancel Myers, and M. Lachmann. Quasispecies made simple. *PLoS Comp. Biol.*, 1:450–460, 2005.

[31] J. J. Bull, R. Sanjuan, and C. O. Wilke. Theory for lethal mutagenesis for viruses. *J. Virology*, 81:2930–2939, 2007.

[32] W. Bussen, S. Raynard, V. Busygina, A. K. Singh, and P. Sung. Holliday junction processing activity of the BLM-TopoIIIα-BLA75 complex. *J. Biol. Chem.*, 282:31484–31492, 2007.

[33] Y. Cao, D. T. Gillespie, and L. R. Petzold. Efficient step size selection for the tau-leaping simulation method. *J. Chem. Phys.*, 124:044109, 2004.

[34] E. A. Carlson. *Mutation. The History of an Idea from Darwin to Genomics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2011.

[35] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.*, 15:1–89, 1943.

[36] K. L. Chung. *Elementary Probability Theory with Stochastic Processes*. Springer-Verlag, New York, 3rd edition, 1979.

[37] A. Constnatinou, X. Chen, C. H. McGowan, and S. C. West. Holliday junction resolution in human cells: Two junction endonucleases with distinct substrate specificities. *EMBO Journal*, 21:5577–5585, 2002.

[38] M. Costanzo, A. Baryshnikova, J. Bellay, Y. Kim, E. D. Spear, C. S. Sevier, H. Ding, J. L. Y. Koh, K. Toufighi, S. Mostafavi, J. Prinz, R. P. S. Onge, B. Van der Sluis, T. Makhnevych, F. J. Vizeacoumar, S. Alizadeh, S. Bahr, R. L. Brost, Y. Chen, M. Cokol, R. Deshpande, Z. Li, Z.-Y. Li, W. Liang, M. Marback, J. Paw, B.-J. San Luis, E. Shuteriqi, A. H. Y. Tong, N. van Dyk, I. M. Wallace, J. A. Whitney, M. T. Weirauch, G. Zhong, H. Zhu, W. A. Houry, M. Brudno, S. Ragibizadeh, B. Papp, C. Pál, F. P. Roth, G. Giaver, C. Nislow, O. G. Troyanskaya, H. Bussey,

G. D. Bader, A.-C. Gingras, Q. D. Morris, P. M. Kim, C. A. Kaiser, C. L.Myers, B. J. Andrews, and C. Boone. The genetic landscape of a cell. *Science*, 327:425–431, 2010.

[39] D. R. Cox and H. D. Miller. *The Theory of Stochastic Processes*. Methuen, London, 1965.

[40] R. T. Cox. *The Algebra of Probable Inference*. The John Hopkins Press, Baltimore, MD, 1961.

[41] J. A. Coyne, N. H. Barton, and M. Turelli. Perspective: A critique of Sewall Wright's shifting balance theory of evolution. *Evolution*, 51:643–671, 1997.

[42] J. A. Coyne, N. H. Barton, and M. Turelli. Is Wright's shifting balance process important in evolution? *Evolution*, 54:306–317, 2000.

[43] J. F. Crow and M. Kimura. *An Introduction to Population Genetics Theory*. Sinauer Associates, Sunderland, MA, 1970.

[44] P. Cull, M. Flahive, and R. Robson. *Difference Equations. From Rabbits to Chaos*. Undergraduate Texts in Mathematics. Springer, New York, 2005.

[45] C. Darwin. *On the Origin of Species by Means of Natural Selection of the or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London, first edition, 1859.

[46] R. Dawkins. *The Selfish Gene*. Oxford Unversity Press, Oxford, UK, 1976.

[47] A. de Candolle. *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.

[48] L. Demetrius. Demographic parameters and natural selection. *Proc. Natl. Acad. Sci. USA*, 71:4645–4647, 1974.

[49] L. Demetrius. Directionality principles in thermodynamics and evolution. *Proc. Natl. Acad. Sci. USA*, 94:3491–498, 1997.

[50] B. Devlin and N. Risch. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics*, 29:311–322, 1995.

[51] E. Domingo, editor. *Quasispecies: Concepts and Implications for Virology*. Springer Verlag, Berlin, 2006.

[52] E. Domingo and J. J. Holland. RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.*, 51:151–178, 1997.

[53] E. Domingo, J. J. Holland, C. Biebriher, and M. Eigen. Quasispecies: The concept and the world. In A. Gibbs, C. Calisher, and F. Garcia-Arenal, editors, *Molecular Evolution of the Viruses*, pages 171–180. Cambridge University Press, Cambridge, UK, 1995.

[54] E. Domingo, C. R. Parrish, and H. John J, editors. *Origin and Evolution of Viruses*. Elsevier, Academic Press, Amsterdam, NL, second edition, 2008.

[55] E. Domingo, D. Szabo, T. Taniguchi, and C. Weissmann. Nucleotide sequence heterogeneity of an RNA phage population. *Cell*, 13:735–744, 1978.

[56] E. Domingo, ed. Virus entry into error catastrophe as a new antiviral strategy. *Virus Research*, 107(2):115–228, 2005.

[57] J. W. Drake. Rates of spontaneous mutation among RNA viruses. *Proc. Natl. Acad. Sci. USA*, 90:4171–4175, 1993.

[58] J. W. Drake, B. Charlesworth, D. Charlesowrth, and J. F. Crow. Rates of spontaneous mutation. *Genetics*, 148:1667–1686, 1998.

[59] D. Duboule. The rise and fall of the Hox gene clusters. *Development*, 134:2549–2560, 2007.

[60] H. J. Dunderdale, F. E. Benson, C. A. Parson, G. J. Sharpless, and S. C. West. Formation and resolution of recombination intermediates by *E. coli* RecA and RuvC proteins. *Nature*, 354:506–510, 1991.

[61] R. A. Dunlap. *The Golden Ratio and Fibonacci Numbers*. World Sceintific, Singapore, 1997.

[62] A. W. F. Edwards. The fundamental theorem of natural selection. *Biological Reviews*, 69:443–474, 1994.

[63] M. Ehrenberg and R. Rigler. Rotational Brownian motion and fluorescence intensity fluctuations. *Chem. Phys.*, 4:390–401, 1974.

[64] P. Ehrenfest and T. Ehrenfest. Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem. *Z. Phys.*, 8:311–314, 1907.

[65] M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58:465–523, 1971.

[66] M. Eigen. On the nature of viral quasispecies. *Trends Microbiol.*, 4:212–214, 1994.

[67] M. Eigen, J. McCaskill, and P. Schuster. Molecular quasispecies. *J. Phys. Chem.*, 92:6881–6891, 1988.

[68] M. Eigen, J. McCaskill, and P. Schuster. The molecular quasispecies. *Adv. Chem. Phys.*, 75:149–263, 1989.

[69] M. Eigen and R. Rigler. Sorting single molecules: Application to diagnpostics and evolutionary biotechnology. *Proc. Netl. Acad. Sci. USA*, 91:5740–5747, 1994.

[70] M. Eigen and P. Schuster. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften*, 64:541–565, 1977.

[71] M. Eigen and P. Schuster. The hypercycle. A principle of natural self-organization. Part B: The abstract hypercycle. *Naturwissenschaften*, 65:7–41, 1978.

[72] M. Eigen and P. Schuster. The hypercycle. A principle of natural self-organization. Part C: The realistic hypercycle. *Naturwissenschaften*, 65:341–369, 1978.

[73] M. Eigen and P. Schuster. Stages of emerging life - Five principles of early organization. *J. Mol. Evol.*, 19:47–61, 1982.

[74] A. Einstein. Über die von der molekular-kinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annal. Phys. (Leipzig)*, 17:549–560, 1905.

[75] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.

[76] T. H. Emigh. A comparison test for Hardy-Weinberg equilibirum. *Biometrics*, 36:627–642, 1980.

[77] H. W. Engl, C. Flamm, P. Kügler, J. Lu, S. Müller, and P. Schuster. Inverse problems in systems biology. *Inverse Problems*, 25:123014, 2009.

[78] P. Erdős and A. Rényi. On random graphs. I. *Publicationes Mathematicae*, 6:290–295, 1959.

[79] P. Erdős and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.

[80] L. Euler. Theoremata arithmetica nova methodo demonstrata. *Novi Comentarii Scientiarum Imperialis Petropolitanæ*, 8:74–104, 1760. Reprinted in Euler's *Commentationes Arithmeticæ Collectæ*, Vol.1, 274-286 and in Euler's *Opera Omnia*, Series 1, Vol.2, 531-555.

[81] L. Euler. *Introductio in Analysin Infinitorum, 1748. English Translation: John Blanton, Introduction to Analysis of the Infinite*, volume I and II. Springer-Verlag, Berlin, 1988.

[82] C. J. Everett and S. Ulam. Multiplicative systems I. *Proc. Natl. Acad. Sci. USA*, 34:403–405, 1948.

[83] C. J. Everett and S. M. Ulam. Multiplicative systems in several variables I. Technical Report LA-683, Los Alamos Scientific Laboratory, 1948.

[84] C. J. Everett and S. M. Ulam. Multiplicative systems in several variables II. Technical Report LA-690, Los Alamos Scientific Laboratory, 1948.

[85] C. J. Everett and S. M. Ulam. Multiplicative systems in several variables III. Technical Report LA-707, Los Alamos Scientific Laboratory, 1948.

[86] W. J. Ewens. *Mathematical Population Genetics*, volume 9 of *Biomathematics Texts*. Springer-Verlag, Berlin, 1979.

[87] W. J. Ewens. An interpretation and proof of the fundamental theorem of natural selection. *Theor. Population Biology*, 36:167–180, 1989.

[88] W. J. Ewens. *Mathematical Population Genetics*. Springer-Verlag, Berlin, second edition, 2004.

[89] W. Feller. Diffusion processes in genetics. In J. Neyman, editor, *Proc. 2nd Berkeley Symp. on Mathematical Statistics and Probability*. University of Caifornia Press, Berkeley, CA, 1951.

[90] J. Felsenstein. The evolutionary advantage of recombination. *Genetics*, 78:737–756, 1974.

[91] R. A. Fisher. The genesis of twins. *Genetics*, 4:489–499, 1919.

[92] R. A. Fisher. *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford, UK, 1930.

[93] R. A. Fisher. Has Mendel's work been rediscovered? *Annals of Science*, pages 115–137, 1936.

[94] D. L. Fisk. Quasi-martingales. *Trans. Amer. Math. Soc.*, 120:369–389, 1965.

[95] C. Flamm, W. Fontana, I. L. Hofacker, and P. Schuster. RNA folding at elementary step resolution. *RNA*, 6:325–338, 2000.

[96] R. A. Flavell, D. L. Szabo, E. F. Bnadle, and C. Weissmann. Site-directed mutagenesis: Effect of an extracistronic mutation on the *in vitro* propagation of bacteriophage Qβ RNA. *Proc. Natl. Acad. Sci. USA*, 72:2170–2174, 1975.

[97] W. Fontana. Modeling Evo-Devo with RNA. *BioSystems*, 24:1164–1177, 2002.

[98] W. Fontana, D. A. M. Konings, P. F. Stadler, and P. Schuster. Statistics of RNA secondary structures. *Biopolymers*, 33:1389–1404, 1993.

[99] W. Fontana, W. Schnabl, and P. Schuster. Physical aspects of evolutionary optimization and adaptation. *Phys. Rev. A*, 40:3301–3321, 1989.

[100] W. Fontana and P. Schuster. A computer model of evolutionary optimization. *Biophys. Chem.*, 26:123–147, 1987.

[101] W. Fontana and P. Schuster. Continuity in evolution. On the nature of transitions. *Science*, 280:1451–1455, 1998.

[102] W. Fontana and P. Schuster. Shaping space. The possible and the attainable in RNA genotype-phenotype mapping. *J. Theor. Biol.*, 194:491–515, 1998.

[103] S. A. Frank and M. Slatkin. Fisher's fundamental theorem ofn natural selection. *TREE*, 7:92–95, 1992.

[104] A. Franklin, A. W. F. Edwards, D. J. Fairbanks, D. L. Hartl, and T. Seidenfeld. *Ending the Mendel-Fisher Controversy*. University of Pittburgh Press, Pittsburgh, PA, 2008.

[105] F. Galton. *Natural Inheritance*. Macmillan & Co., London, second american edition, 1889. App. F, pp.241-248.

[106] J. Garcia-Fernàndez. The genesis and evolution of homeobox gene clusters. *Nature Reviews Genetics*, 6:881–892, 2005.

[107] C. W. Gardiner. *Stochastic Methods. A Handbook for the Natural Sciences and Social Sciences*. Springer Series in Synergetics. Springer-Verlag, Berlin, fourth edition, 2009.

[108] T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic togle switch in *escherichia coli*. *Nature*, 403:339–342, 2000.

[109] S. Gavrilets. Evolution and speciation on holey adaptive landscapes. *Trends in Ecology and Evolution*, 12:307–312, 1997.

[110] A. Gierer and H. Meinhardt. A theory of biologicl pattern formation. *Kybernetik*, 12:30–39, 1972.

[111] S. F. Gilbert. *Developmental Biology*. Sinauer Associates, Sunderland, MA, sixth edition, 2000.

[112] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, 22:403–434, 1976.

[113] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.

[114] D. T. Gillespie. *Markov Processes: An Introduction for Physical Scientists*. Academic Press, San Diego, CA, 1992.

[115] D. T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A*, 188:404–425, 1992.

[116] D. T. Gillespie. Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral. *Phys. Rev. E*, 54:2084–2091, 1996.

[117] D. T. Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, 58:35–55, 2007.

[118] N. S. Goel and N. Richter-Dyn. *Stochastic Models in Biology*. Academic Press, New York, 1974.

[119] P. Gönczy and L. S. Rose. Asymmetric cell division and axis formation in the embryo. In *The C. elegans Research Community*, editor, *WormBook. 1.30.1.* `http://www.wormbook.org`, WormBook, 2005.

[120] C. J. Goodnight and M. J. Wade. The ongoing synthesis: A reply to Coyne, Barton and Turelli. *Evolution*, 54:317–324, 2000.

[121] M. Gösch and R. Rigler. Fluorescence correlation spectroscopy of molecular motions and kinetics. *Advanced Drug Delivery Reviews*, 57:169–190, 2005.

[122] S. Gottesman. The small RNA regulators of *escherichia coli*: Roles and mechanisms. *Annu. Rev. Microbiol.*, 58:303–328, 2004.

[123] I. S. Gradstein and I. M. Ryshik. *Tables of Series, Products, and Integrals*, volume 1. Verlag Harri Deutsch, Thun, DE, 1981. In German and English. Translated from Russian by Ludwig Boll, Berlin.

[124] R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics*. Addison-Wesley Publishing Co., Reading, MA, second edition, 1994.

[125] A. J. F. Griffiths, J. H. Miller, D. T. Suzuki, R. C. Lewontin, and W. M. Gelbart. *An Introduction to Genetic Analysis*. W. H. Freeman, New York, seventh edition, 2000.

[126] W. Grüner, R. Giegerich, D. Strothmann, C. Reidys, J. Weber, I. L. Hofacker, and P. Schuster. Analysis of RNA sequence structure maps by exhaustive enumeration. I. Neutral networks. *Mh. Chem.*, 127:355–374, 1996.

[127] W. Grüner, R. Giegerich, D. Strothmann, C. Reidys, J. Weber, I. L. Hofacker, and P. Schuster. Analysis of RNA sequence structure maps by exhaustive enumeration. II. Structures of neutral networks and shape space covering. *Mh. Chem.*, 127:375–389, 1996.

[128] M. Güell, V. van Noort, E. Yus, W.-H. Chen, J. Leigh-Bell, K. Michalodimitrakis, T. Yamada, M. Arumugam, T. Doerks, S. Kühner, M. Rode, M. Suyama, S. Schmidt, A.-C. Gavin, P. Bork, and L. Serrano. Transcriptome complexity in a genome-reduced bacterium. *Science*, 326:1268–1271, 2009.

[129] J. B. S. Haldane. An exact test for randomness of mating. *J. Genetics*, 52:631–635, 1954.

[130] M. B. Hamilton. *Population Genetics*. Wiley-Blackwell, Oxford, UK, 2009.

[131] W. D. Hamilton. The genetical evolution of social behaviour. i. *J. Theor. Biol.*, 7:1–16, 1964.

[132] W. D. Hamilton. The genetical evolution of social behaviour. ii. *J. Theor. Biol.*, 7:17–52, 1964.

[133] R. W. Hamming. Error detecting and error correcting codes. *Bell Syst. Tech. J.*, 29:147–160, 1950.

[134] G. H. Hardy. Mendelian proportions in a mixed population. *Science*, 28:49–50, 1908.

[135] T. E. Harris. *The Theory of Branching Processes*. Dover Publications, New York, 1989.

[136] D. L. Hartl and A. G. Clark. *Principles of Population Genetics*. Sinauer Associates, Sunderland, MA, third edition, 1997.

[137] E. L. Haseltine and F. H. Arnold. Synthetic gene circuits: Design with directed evolution. *Annu. Rev. Biophys. Biomol. Struct.*, 36:1–19, 2007.

[138] D. Hawkins and S. Ulam. Theory of multiplicative processes I. Technical Report LADC-265, Los Alamos Scientific Laboratory, 1944.

[139] N. Hawkins and G. Garriga. Asymmetric cell division: From A to Z. *Genes & Development*, 12:3625–3638, 1998.

[140] Y. Hayashi, T. Aita, H. Toyota, Y. Husimi, I. Urabe, and T. Yomo. Experimental rugged fiteness landscape in protein sequence space. *PLoS One*, 1:e96, 2006.

[141] C. R. Heathcote and J. E. Moyal. The random walk (in continuous time) and its application to the theory of queues. *Biometrika*, 46:400–411, 1959.

[142] C. C. Heyde and E. Seneta. 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.

[143] M. W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, New York, 1974.

[144] M. W. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and an Introduction to Chaos*. Elsevier, Amsterdam, second edition, 2004.

[145] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Mh. Chemie*, 125:167–188, 1994.

[146] I. L. Hofacker, P. Schuster, and P. F. Stadler. Combinatorics of RNA secondary structures. *Discr. Appl. Math.*, 89:177–207, 1998.

[147] J. Hofbauer and K. Sigmund. *Dynamical Systems and the Theory of Evolution*. Cambridge University Press, Cambridge, UK, 1988.

[148] J. Hohlbein, K. Gryte, M. Heilemann, and A. N. Kapanidis. Surfing on a new wave of single-molecule fluourescence methods. *Phys. Biol.*, 7:031001, 2010.

[149] J. J. Holland, J. C. de la Torre, and D. Steinhauer. RNA virus populations as quasispecies. *Curr. Topics Microbiol. Immunol.*, 176:1–20, 1992.

[150] R. Holliday. A mechanism for gene conversion in fungi. *Genetical Research*, 5:282–304, 1964.

[151] N. M. Hollingsworth and B. Steven J. The Mus81 solution to resolution: Generating meiotic crossovers without Holliday junctions. *Genes Dev.*, 18:117–125, 2004.

[152] S. P. Hubbel. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton, NJ, 2001.

[153] Y. Husimi. Selection and evolution of a bacteriophages in cellstat. *Adv. Biophys.*, 25:1–43, 1989.

[154] Y. Husimi, K. Nishigaki, Y. Kinoshita, and T. Tanaka. Cellstat – A continuous culture system of a bacteriophage for the study of the mutation rate and the selection process at the DNA level. *Rev. Sci. Instrum.*, 53:517–522, 1982.

[155] C. A. Hutchinson III, S. Phillips, M. H. Edgell, S. Gillam, P. Jahnke, and M. Smith. Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.*, 25:6551–6560, 1978.

[156] M. A. Huynen, P. F. Stadler, and W. Fontana. Smoothness within ruggedness. The role of neutrality in adaptation. *Proc. Natl. Acad. Sci. USA*, 93:397–401, 1996.

[157] S. C. Y. Ip, U. Rass, M. G. Blanco, H. R. Flynn, J. M. Skehel, and S. C. West. Identification of Holliday junction resolvases from humans and yeast. *Nature*, 456:357–361, 2008.

[158] K. Ishida. Stochastic model for bimolecular reaction. *J. Chem. Phys.*, 41:2472–2478, 1964.

[159] K. Itō. Stochastic integral. *Proc. Imp. Acad. Tokyo*, 20:519–524, 1944.

[160] K. Itō. On stochastic differential equations. *Mem. Amer. Math. Soc.*, 4:1–51, 1951.

[161] C. Jäckel, P. Kast, and D. Hilvert. Protein design by directed evolution. *Annu. Rev. Biophys.*, 37:153–173, 2008.

[162] C. James J. DNA topoisomerases: Structure, function, and mechanism. *Annu. Rev. Biochem.*, 70:369–413, 2001.

[163] A. Janshoff, M. Neitzert, Y. Oberdörfer, and H. Fuchs. Force spectroscopy of molecular systems – single molecule spectroscopy of polymers and biomolecules. *Angew. Chem. Int. Ed.*, 39:3212–3237, 2000.

[164] E. T. Jaynes. *Probability Theory. The Logic of Science*. Cambridge University Press, Cambridge, UK, 2003.

[165] B. L. Jones, R. H. Enns, and S. S. Rangnekar. On the theory of selection of coupled macromolecular systems. *Bull. Math. Biol.*, 38:15–28, 1976.

[166] G. F. Joyce. Directed evolution of nucleic acid enzymes. *Annu. Rev. Biochem.*, 73:791–836, 2004.

[167] H. Judson. *The Eighth Day of Creation. The Makers of the Revolution in Biology*. Jonathan Cape, London, 1979.

[168] M. Kallis, B. W. Birren, and E. S. Lander. Proof and evolutionarty analysis of ancient genome duplication in the yeast *saccharomyces cerevisiae*. *Nature*, 428:617–624, 2004.

[169] S. Kauffman and S. Levin. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.*, 128:11–45, 1987.

[170] S. A. Kauffman and E. D. Weinberger. The N-k model of rugged fitness landscapes and its application to the maturation of the immune response. *J. Theor. Biol.*, 141:211–245, 1989.

[171] D. G. Kendall. Branching processes since 1873. *J. of the London Mathematical Society*, 41:386–406, 1966.

[172] D. G. Kendall. The genalogy of genealogy: Branching processes before (an after) 1873. *Bull. of the London Mathematical Society*, 7:225–253, 1975.

[173] M. Kimura. Evolutionary rate at the molecular level. *Nature*, 217:624–626, 1968.

[174] M. Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK, 1983.

[175] J. L. King and T. H. Jukes. Non-darwinian evolution. *Science*, 164:788–798, 1969.

[176] S. Klussmann, editor. *The Aptamer Handbook. Functional Oligonucleotides and Their Applications*. Wiley-VCh Verlag, Weinheim, DE, 2006.

[177] J. Knoblich. Mechnanisms of asymmetric stem cell division. *Cell*, 132:583–597, 2008.

[178] D. E. Knuth. *The Art of Computer Programming. Vol.I: Fundamental Algorithms*. Addison-Wesley Publishing Co., Reading, MA, third edition, 1997.

[179] A. N. Kolmogorov and N. A. Dmitriev. "Zur Lösung einer biologischen Aufgabe". *Isvestiya Nauchno-Issledovatel'skogo Instituta Matematiki i Mekhaniki pri Tomskom Gosudarstvennom Universitete*, 2:1–12, 1938.

[180] A. N. Kolmogorov and N. A. Dmitriev. Branching stochastic processes. *Doklady Akad. Nauk U.S.S.R.*, 56:5–8, 1947.

[181] R. D. Kouyos, G. E. Leventhal, T. Hinkley, M. Haddad, J. M. Whitcomb, C. J. Petropoulos, and S. Bonhoeffer. Exploring the complexity of the HIV-1 fitness landscape. *PLoS Genetics*, 8:e1002551, 2012.

[182] S. Kühner, V. van Noort, M. J. Betts, A. Leo-Macias, C. Batisse, M. Rode, T. Yamada, T. Maier, S. Bader, P. Beltran-Alvarez, D. Castaño-Diez, W.-H. Chen, D. Devos, M. Güell, T. Norambuena, I. Racke, V. Rybin, A. Schmidt, E. Yus, R. Aebersold, R. Herrmann, B. Böttcher, A. S. Frangakis, R. B. Russell, L. Serrano, P. Bork, and A.-C. Gavin. Proteome organization in a genome-reduced bacterium. *Science*, 326:1235–1240, 2009.

[183] S. Kumar. Molecular clocks: Four decades of evolution. *Nature Reviews Genetics*, 6:654–662, 2005.

[184] T. A. Kunkel. Rapid an efficient site-specific mutagenesis wihtout phenotypic selectiion. *Proc. Natl. Acad. Sci. USA*, 82:488–492, 1985.

[185] S. Kuraku and A. Meyer. The evolution and maintenance of Hox gene clusters in vertebrates and the teleost-specific genome duplication. *Internat. J. Dev. Biol.*, 53:765–773, 2009.

[186] P. Langevin. Sur la théorie du mouvement Brownien. *Comptes Rendues*, 146:530–533, 1908.

[187] I. Leuthäusser. An exact correspondence between Eigen's evolution model and a two-dimensional ising system. *J. Chem. Phys.*, 84:1884–1885, 1986.

[188] I. Leuthäusser. Statistical mechanics of Eigen's evolution model. *J. Stat. Phys.*, 48:343–360, 1987.

[189] S. Lifson. On the theory of helix-coil transitions in polypeptides. *J. Chem. Phys.*, 34:1963–1974, 1961.

[190] D. Magde, E. Elson, and W. W. Webb. Thermodynamic fluctuations in a reating system – Measurement by flourescence correlation spectroscopy. *Phys. Rev. Letters*, 29:705–708, 1972.

[191] B. W. J. Mahy and M. H. V. van Regenmortel, editors. *Desk Encyclopedia of General Virology*. Elsevier, Academic Press, Oxford, UK, 2010.

[192] P. K. Maini, K. J. Painter, and H. Nguyen Phong Chau. Spatial pattern formation in chemical and biological systems. *J. Chem. Soc., Faraday Trans.*, 93:3601–3610, 1997.

[193] T. R. Malthus. *An Essay of the Principle of Population as it Affects the Future Improvement of Society*. J. Johnson, London, 1798.

[194] M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, and R. R. Breaker. Riboswitches control fundamental biochemical pathways in *bacillus subtilis* and other bacteria. *Cell*, 113:577–596, 2003.

[195] M. Mann and K. Klemm. Efficient exploration of discrete energy landscapes. *Phys. Rev. E*, 83:011113, 2011.

[196] M. A. Marchisio and J. Stelling. Computational design of synthetic gene circuit with composable parts. *Bioinformatics*, 24:1903–1910, 2008.

[197] G. H. Markx, C. L. Davey, and D. B. Kell. The permittistat – A novel type of turbidostat. *J. General Microbiol.*, 137:735–743, 1991.

[198] T. Maruyama. *Stochastic Problems in Population Genetics*. Springer-Verlag, Berlin, 1977.

[199] D. H. Mathews, M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, and D. H. Turner. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA*, 101:7287–7292, 2004.

[200] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, 288:911–940, 1999.

[201] J. Maynard-Smith. Natural selection and the concept of a protein space. *Nature*, 225:563–564, 1970.

[202] E. Mayr. *The Growth of Biological Thought. Diversity, Evolution, and Inhertiance*. The Belknap Press of Harvard University Press, 1982.

[203] D. A. McQuarrie, C. J. Jachimowski, and M. E. Russell. Kinetics of small systems. II. *J. Chem. Phys.*, 40:2914–2921, 1964.

[204] H. Meinhardt. *Models of Biological Pattern Formation*. Academic Press, London, 1982.

[205] H. Meinhardt and A. Gierer. Pattern formation by local self-activation and lateral inhibition. *BioEssays*, 22:753–760, 2000.

[206] G. Mendel. Versuche über Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereins in Brünn*, IV:3–47, 1866. In German.

[207] G. Mendel. Über einige aus künstlicher Befruchtung gewonnenen Hieracium-Bastarde. *Verhandlungen des naturforschenden Vereins in Brünn*, VIII:26–31, 1870. In German.

[208] M. S. Meselson and C. M. Radding. A general model for genetic recombination. *Proc. Natl. Acad. Sci. USA*, 72:358–361, 1975.

[209] A. Messiah. *Quantum Mechanics*, volume II. North-Holland Publishing Company, Amsterdam, NL, 1970. Translated from the French by J. Potter.

[210] E. W. Montroll. Stochastic processes and chemical kinetics. In W. M. Muller, editor, *Energetics in Metallurgical Phenomenon*, volume 3, pages 123–187. Gordon & Breach, New York, 1967.

[211] E. W. Montroll and K. E. Shuler. The application of the theory of stochastic processes to chemical kinetics. *Adv. Chem. Phys.*, 1:361–399, 1958.

[212] P. A. P. Moran. Random processes on genetics. *Proc. Camb. Phil. Soc.*, 54:60–71, 1958.

[213] P. A. P. Moran. *The Statistical Processes of Evolutionary Theroy*. Clarendon Press, Oxford, UK, 1962.

[214] P. A. P. Moran. On the nonexistence of adaptive topographies. *Ann. Hum. Genet.*, 27:383–393, 1964.

[215] D. W. Mount. *Bioinformatics. Sequence and Genome Analysis*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, second edition, 2004.

[216] A. E. Mourant, A. Kopec, and K. Domaniewska-Sobczak. *The Distribution of Human Blood Groups and Other Polymorphisms*. Oxford University Press, New York, second edition, 1976.

[217] S. Muñoz-Galván, C. Tous, M. G. Blanco, E. K. Schwartz, K. T. Emsen, S. C. West, Wolf-Dietrich, and A. Aguilera. Distinct roles of Mus81, Yen1, Slx1-Slx4, and Rad1 nucleases in the repair of replication-born double-strand breaks by sister chromatid exchange. *Molecular and Cellular Biology*, 32:1592–1603, 2012.

[218] H. J. Muller. Variation due to change in the individual gene. *American Naturalist*, 56:32–50, 1922.

[219] H. J. Muller. Artificial transmutation of the gene. *Science*, 66:84–87, 1927.

[220] H. J. Muller. Some genetic aspects of sex. *American Naturalist*, 66:118–138, 1932.

[221] H. J. Muller. The relation of recombination to mutational advance. *Mutat. Res.*, 106:2–9, 1964.

[222] K. Nasmyth. Disseminating the genome: Joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu. Rev. Genet.*, 35:673–745, 2001.

[223] G. Nicolis and I. Prigogine. *Self-Organization in Nonequilibrium Systems*. John Wiley & Sons, New York, 1977.

[224] M. Nowak and P. Schuster. Error thresholds of replication in finite populations. Mutation frequencies and the onset of Muller's ratchet. *J. Theor. Biol.*, 137:375–395, 1989.

[225] M. Nowak, C. E. Tarnita, and E. O. Wilson. The evolution of eusociality. *Nature*, 466:1057–1062, 2010.

[226] S. Ohno. *Evolution by Gene Duplication*. Springer-Verlag, New York, 1970.

[227] T. Ohta. Mechanisms of molecular evolution. *Phil. Trans. Roy. Soc. London B*, 355:1623–1626, 2000.

[228] S. Okasha. Fisher's fundamental theorem of natural selection – A philosophical analysis. *Brit. J. Phil. Sci.*, 59:319–351, 2008.

[229] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes. Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.*, 48:545–600, 1997.

[230] M. Oram, A. Keeley, and I. Tsaneva. Holliday junction resolvase in *schizosaccharomyces pombe* has identical endonuclease activity to the CCE1 homologue YDC2. *Nucleic Acids Research*, 26:594–601, 1998.

[231] T. L. Orr-Weaver and J. W. Szostak. Yeast recombination: The association between double-strand gap repair and crossing-over. *Proc. Natl. Acad. Sci. USA*, 80:4417–4421, 1983.

[232] R. D. M. Page and E. C. Holmes. *Molecular Evolution. A Phylogenetic Approach*. Blackwell Science Ltd., Oxford, UK, 1998.

[233] K. Pearson. 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.

[234] P. E. Phillipson and P. Schuster. *Modeling by Nonlinear Differential Equations. Dissipative and Conservative Processes*, volume 69 of *World Scientific Series on Nonlinear Science A*. World Scientific, Singapore, 2009.

[235] J. N. Pitt and A. R. Ferré-D'Amaré. Rapid construction of empirical RNA fitness landscapes. *Science*, 330:376–379, 2010.

[236] G. R. Price. Fisher's 'fundamental theorem' made clear. *Annals of Human Genetics*, 36:129–140, 1972.

[237] J. S. Reader and G. F. Joyce. A ribozyme composed of only two different nucleotides. *Nature*, 420:841–844, 2002.

[238] C. Reidys, P. F. Stadler, and P. Schuster. Generic properties of combinatory maps. Neutral networks of RNA secondary structure. *Bull. Math. Biol.*, 59:339–397, 1997.

[239] C. M. Reidys and P. F. Stadler. Combinatorial landscapes. *SIAM Review*, 44:3–54, 2002.

[240] R. Rigler. Fluorescence correlation, single molecule detection and large number screening. Applications in biotechnology. *J. Biotechnology*, 41:177–186, 1995.

[241] R. W. Robinett. *Quantum Mechanics. Classical Results, Modern Systems, and Visualized Examples*. Oxford University Press, New York, 1997.

[242] F. Schlögl. Chemical reaction models for non-equilibrium phase transitions. *Z. Physik*, 253:147–161, 1972.

[243] W. Schnabl, P. F. Stadler, C. Forst, and P. Schuster. Full characterizatin of a strang attractoe. Chaotic dynamics on low-dimensional replicator systems. *Physica D*, 48:65–90, 1991.

[244] M. Schubert and G. Weber. *Quantentheorie. Grundlagen und Anwendungen*. Spektrum Akademischer Verlag, Heidelberg, DE, 1993. In German.

[245] P. Schuster. Potential functions and molecular evolution. In M. Markus, S. C. Müller, and G. Nicolis, editors, *From Chemical to Biological Organization. Springer Series in Synergetics*, volume 39, pages 149–165. Springer-Verlag, Berlin, 1988.

[246] P. Schuster. How does complexity arise in evolution. Nature's recipe for mastering scarcity, abundance, and unpredictability. *Complexity*, 2(1):22–30, 1996.

[247] P. Schuster. Molecular insight into the evolution of phenotypes. In J. P. Crutchfield and P. Schuster, editors, *Evolutionary Dynamics – Exploring the Interplay of Accident, Selection, Neutrality, and Function*, pages 163–215. Oxford University Press, New York, 2003.

[248] P. Schuster. Prediction of RNA secondary structures: From theory to models and real molecules. *Reports on Progress in Physics*, 69:1419–1477, 2006.

[249] P. Schuster. Is there a Newton of the blade of grass? The complex relation between mathematics, physics, and biology. *Complexity*, 16(6):5–9, 2011.

[250] P. Schuster. Mathematical modeling of evolution. Solved and open problems. *Theory in Biosciences*, 130:71–89, 2011.

[251] P. Schuster. Stochastic chemical kinetics. A special course on probability and stochastic processes for physicists, chemiststs, and biologists, University of Vienna, Download: `www.tbi.univie.ac.at/~pks/Preprints/stochast.pdf`, 2011.

[252] P. Schuster, W. Fontana, P. F. Stadler, and I. L. Hofacker. From sequences to shapes and back: A case study in RNA secondary structures. *Proc. Roy. Soc. Lond. B*, 255:279–284, 1994.

[253] P. Schuster and K. Sigmund. Replicator dynamics. *J. Theor. Biol.*, 100:533–538, 1983.

[254] P. Schuster and K. Sigmund. Random selection - A simple model based on linear birth and death processes. *Bull. Math. Biol.*, 46:11–17, 1984.

[255] P. Schuster and K. Sigmund. Dynamics of evolutionary optimization. *Ber. Bunsenges. Phys. Chem.*, 89:668–682, 1985.

[256] P. Schuster and J. Swetina. Stationary mutant distribution and evolutionary optimization. *Bull. Math. Biol.*, 50:635–660, 1988.

[257] J. Sekiguchi, C. Cheng, and S. Shuman. Resolution of a Holliday junction by vaccinia topoisomerase requires a spacer DNA segment 3' of the **CCCTT**↓ cleavage sites. *Nucleic Acids Research*, 28:2658–2663, 2000.

[258] J. Sekiguchi, N. C. Seeman, and S. Shuman. Resolution of Holliday junctions by eukaryotic DNA topoisomerase I. *Proc. Natl. Acad. Sci.*, 93:785–789, 1996.

[259] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer-Verlag, New York, second edition, 1981.

[260] A. Serganov and D. J. Patel. Ribozymes, riboswitches and beyond: Regulation of gene expression without proteins. *Nature Reviews Genetics*, 8:776–790, 2007.

[261] S. Shahshahani. A new mathematical framework for the study of linkage and selection. *Mem. Am. Math. Soc.*, 211, 1979.

[262] K. E. Shuler, G. H. Weiss, and K. Anderson. Studies in nonequilibrium rate processes. V. The relaxation of moments derived from a master equation. *J. Math. Phys.*, 3:550–556, 1962.

[263] H. W. Siemens. *Die Zwillingspathologie. Ihre Bedeutung, ihre Methodik, ihre bisherigen Ergebnisse*. J. Springer, Berlin, 1924.

[264] L. E. Sigler. *Fibonacci's Liber Abaci: A Translation into Modern English of Leonardo Pisano's Book of Calculation*. Springer-Verlag, New York, 2002. A translation of Leonardo Pisano's book of 1202 from Latin into modern English.

[265] P. Singh. The so-called Fibonacci numbers in ancient and medieval India. *Historia Mathematica*, 12:229–244, 1985.

[266] R. A. Skipper Jr. The persistence of the R.A. Fisher – Sewall Wright controversy. *Biology and Philosophy*, 17:341–367, 2002.

[267] R. V. Solé, S. C. Manrubia, B. Luque, J. Delgado, and J. Bascompte. Phase transitions and complex systems – Simple, nonlinear models capture complex systems at the edge of chaos. *Complexity*, 1(4):13–26, 1996.

[268] S. Spiegelman. An approach to the experimental analysis of precellular evolution. *Quart. Rev. Biophys.*, 4:213–253, 1971.

[269] B. R. M. Stadler, P. F. Stadler, M. Shpak, and G. P. Wagner. Recombination spaces, metrics, and pretopologies. *Z. Phys. Chem.*, 216:217–234, 2002.

[270] B. R. M. Stadler, P. F. Stadler, G. P. Wagner, and W. Fontana. The topology of the possible: Formal spaces underlying patterns of evolutionary change. *J. Theor. Biol.*, 213:241–274, 2001.

[271] P. F. Stadler. Landscapes and their correlation functions. *J. Math. Chem.*, 20:1–45, 1996.

[272] P. F. Stadler and G. P. Wagner. Algebraic theory of recombination spaces. *Evolutionary Computation*, 5:241–275, 1998.

[273] F. W. Stahl. The Holliday junction on its thirtieth anniversary. *Genetics*, 138:241–246, 1995.

[274] F. W. Stahl, I. Kobashi, and M. M. Stahl. In phage λ, *cos* is a recombinator in the red pathway. *J. Mol. Biol.*, 181:199–209, 1985.

[275] J. F. Steffensen. "deux problème du calcul des probabilités". *Ann. Inst. Henri Poincaré*, 3:319–344, 1933.

[276] B. Stillman, D. Stewart, and J. Witkowski, editors. *Evolution. The Molecular Landscape*, volume LXXIV of *Cold Spring Harbor Symposia on Quantitative Biology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2009.

[277] R. L. Stratonovich. *Introduction to the Theory of Random Noise*. Gordon and Breach, New York, 1963.

[278] I.-M. Sun, R. Wladerkiewicz, and A. Riuz-Carrillo. Histone H5 in the control of DNA synthesis and cell proliferation. *Science*, 245:68–71, 1989.

[279] J. M. Svendsen and J. W. Harper. GEN1/Yen1 and the SLX4 complex: Solution to the problem of Holliday junction resolution. *Genes Dev.*, 24:521–536, 2010.

[280] J. Swetina and P. Schuster. Self-replication with errors - A model for polynucleotide replication. *Biophys. Chem.*, 16:329–345, 1982.

[281] L. S. Symington and W. K. Holloman. Resolving resolvases: The final act? *Molecular Cell*, 32:603–604, 2008.

[282] P. Tarazona. Error thresholds for molecular quasispecies as phase transitions: From simple landscapes to spin glasses. *Phys. Rev. A*, 45:6038–6050, 1992.

[283] H. Tejero, A. Marín, and F. Moran. Effect of lethality on the extinction and on the error threshold of quasispecies. *J. Theor. Biol.*, 262:733–741, 2010.

[284] D. S. Thaler, M. M. Stahl, and F. W. Stahl. Test of the double-strand-break repair model for red-mediated recombination of phage λ and plasmid λdv. *Genetics*, 116:501–511, 1987.

[285] C. J. Thompson and J. L. McBride. On Eigen's theory of the self-organization of matter and the evolution of biological macromolecules. *Math. Biosci.*, 21:127–142, 1974.

[286] R. C. Tolman. *The Principle of Statistical Mechanics*. Oxford University Press, Oxford, UK, 1938.

[287] A. M. Turing. The chemical basis of morphogenesis. *Phil. Trans. Roy. Soc. London B*, 237(641):37–72, 1952.

[288] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the Brownian motion. *Phys. Rev.*, 36:823–841, 1930.

[289] Y. Van de Peer, S. Maere, and A. Meyer. The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics*, 10:725–732, 2009.

[290] T. van den Berg. *Calibrating the Ornstein-Uhlenbeck-Vasicek model*. Sitmo – Custom Financial Research and Development Services, `www.sitmo.com/article/calibrating-the-ornstein-uhlenbeck-model/`, May 2011.

[291] O. Vasicek. An equlibrium characterization of the term structure. *J. Financial Economics*, 5:177–188, 1977.

[292] P. Verhulst. Notice sur la loi que la population pursuit dans son accroisement. *Corresp. Math. Phys.*, 10:113–121, 1838.

[293] P. Verhulst. Recherches mathématiques sur la loi d'accroisement de la population. *Nouv. Mém. de l'Academie Royale des Sci. et Belles-Lettres de Bruxelles*, 18:1–41, 1845.

[294] P. Verhulst. Deuxième mémoire sur la loi d'accroisement de la population. *Mém. de l'Academie Royale des Sci., des Lettres et de Beaux-Arts de Belgique*, 20:1–32, 1847.

[295] G. von Kiedrowski, B. Wlotzka, J. Helbig, M. Matzen, and S. Jordan. Parabolic growth of a self-replicating hexanucleotide bearing a 3'-5'-phosphoamidate linkage. *Angew. Chem. Internat. Ed.*, 30:423–426, 1991.

[296] M. von Smoluchowski. Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Annal. Phys. (Leipzig)*, 21:756–780, 1906.

[297] M. J. Wade and C. J. Goodnight. Perspective: The theories of Fisher and Wright in the context of metapopulations: When nature does many small experiments. *Evolution*, 52:1537–1548, 1998.

[298] N. Wagner, E. Tannenbaum, and G. Ashkenasy. Second order catalytic quasispecies yields discontinuous mean fitness at error threshold. *Phys. Rev. Letters*, 104:188101, 2010.

[299] S. Wahlund. Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas*, 11:65–106, 1928.

[300] M. S. Waterman. Secondary structures of single stranded nucleic acids. *Adv. Math. Suppl. Studies*, I:167–212, 1978.

[301] H. W. Watson and F. Galton. On the probability of the extinction of families. *J. Anthropological Institute of Great Britain and Ireland*, 4:138–144, 1875.

[302] J. D. Watson and F. H. C. Crick. A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[303] W. Weinberg. Über den NAchweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg*, 64:369–382, 1908.

[304] A. Weismann. *Essays Upon Heridity*, volume 1 and 2. Clarendon Press, Oxford, UK, electronic scholarly publishing edition, 1889.

[305] A. Weismann. *Das Keimplasma. Eine Theorie der Vererbung.* Fischer, Jena, DE, 1892.

[306] J. J. Welch and D. Waxman. The nk model and population genetics. *J. Theor. Biol.*, 234:329–340, 2005.

[307] M. C. Whitby. Making crossover during meiosis. *Biochem. Soc. Trans.*, 33:1451–1455, 2005.

[308] T. Wiehe. Model dependency of error thresholds: The role of fitness functions and contrasts between the finite and infinite sites models. *Genet. Res. Camb.*, 69:127–136, 1997.

[309] C. O. Wilke, J. L. Wang, and C. Ofria. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, 412:331–333, 2001.

[310] D. Williams. *Diffusions, Markov Processes and Martingales. Volume 1: Foundations*. John Wiley & Sons, Chichester, UK, 1979.

[311] P. R. Wills, S. A. Kauffman, B. M. R. Stadler, and P. F. Stadler. Selection dynamics in autocatalytic systems: Templates replicating through binary ligation. *Bull. Math. Biol.*, 60:1073–1098, 1998.

[312] W. C. Winkler. Metabolic monitoring by bacterial mRNAs. *Arch. Microbiol.*, 183:151–159, 2005.

[313] M. T. Wolfinger, W. A. Svrcek-Seiler, C. Flamm, I. L. Hofacker, and P. F. Stadler. Efficient computation of RNA folding dynamics. *J. Phys. A: Mathematical and General*, 37:4731–4741, 2004.

[314] P. G. Wolynes. Energy landscapes and solved protein-folding problems. *Phil. Trans. Roy. Soc. A*, 363:453–467, 1997.

[315] M. B. Wreissman. Fluctuation spectroscopy. *Annu. Rrev. Phys. Chem.*, 32:205–232, 1981.

[316] S. J. Wrenn and P. B. Harbury. Chemical evolution as a tool for molecular discovery. *Annu. Rev. Biochem.*, 76:331–349, 2007.

[317] S. Wright. Coefficients of inbreeding and relationship. *American Naturalist*, 56:330–338, 1922.

[318] S. Wright. Evolution in Mendelian populations. *Genetics*, 16:97–159, 1931.

[319] S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In D. F. Jones, editor, *Int. Proceedings of the Sixth International Congress on Genetics*, volume 1, pages 356–366, Ithaca, NY, 1932. Brooklyn Botanic Garden.

[320] L. Wu and I. D.Hickson. The Bloom's syndrome helicase sipresses crossing over during homologous recombination. *Nature*, 426:870–874, 2003.

[321] K. Young and J. P. Crutchfield. Fluctuation spectroscopy. *Chaos, Solitons & Fractals*, 4:5–39, 1994.

[322] E. Yus, T. Maier, K. Michalodimitrakis, V. van Noort, T. Yamada, W.-H. Chen, J. A. Wodke, M. Güell, S. Martínez, R. Bourgeois, S. Kühner, E. Raineri, I. Letunic, O. V. Kalinina, M. Rode, R. Herrmann, R. Gutiérez-Gallego, R. B. Russell, A.-C. Gavin, P. Bork, and L. Serrano. Impact of genome reduction on bacterial metabolism and its regulation. *Science*, 326:1263–1268, 2009.

[323] W. K. Zhang and X. Zhang. Single molecule mechanochemistry of macromolecules. *Prog. Polym. Sci.*, 28:1271–1295, 2003.

[324] B. H. Zimm. Theory of "melting" of the helical form in double chains of the DNA type. *J. Chem. Phys.*, 33:1349–1356, 1960.

[325] B. H. Zimm and J. K. Bragg. Theory of the phase transition between helix and coil in polypeptide chains. *J. Chem. Phys.*, 31:526–535, 1959.

[326] M. Zuker. On finding all suboptimal foldings of an RNA molecule. *Science*, 244:48–52, 1989.

[327] M. Zuker. The use of dynamic programming algorithms in RNA secondary structure prediction. In M. S. Waterman, editor, *Mathematical Methods for DNA Sequences*, pages 159–184. CRC Press, Boca Raton (FL), 1989.

[328] M. Zuker and P. Stiegler. Optimal computer folding of larger RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9:133–148, 1981.

[329] D. Zwillinger. *Handbook of Differential Equations*. Academic Press, San Diego, CA, third edition, 1998.

Contents

Index