

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS
DEPARTAMENTO DE FÍSICA E INFORMÁTICA

Modelos de Integração de Informação em Evolução Pré-Biótica

Paulo Roberto de Araújo Campos

*Tese apresentada ao Instituto de Física de
São Carlos, Universidade de São Paulo, pa-
ra obtenção do Título de Doutor em Ciências:
Física Básica*

ORIENTADOR: *Prof. Dr. José Fernando Fontanari*

SÃO CARLOS

2001

Agradecimentos

Agradeço em primeiro lugar à Deus, pelas oportunidades e apoio em todas as circunstâncias.

Agradecimento em especial ao Prof. Dr. José Fernando Fontanari, pela orientação, paciência e pelo meu aprendizado durante esses três anos e meio de convivência. Aproveito a oportunidade para expressar minha admiração pela sua competência e seriedade profissional.

Ao Prof. Dr. Peter F. Stadler, com quem desenvolvemos o estudo de hiperciclos, e pela sua hospitalidade durante meu estágio de 3 meses à Universidade de Viena.

Ao Prof. Dr. Christoph Adami e Dr. Claus Wilke, com quem colaborei e desenvolvi o estudo de relevos rugosos, e também pela receptividade durante a minha estada de 2 meses no California Institute of Technology (CALTECH).

À minha esposa, por todo amor, carinho e dedicação em todos os instantes; pela paciência e apoio tanto no trabalho como em outras atividades.

À minha mãe, pelo apoio e dedicação que nunca me faltaram e pelo seu esforço e sacrifício realizados, de forma a permitir uma boa educação e crescimento profissional de seus filhos.

Ao meu irmão, por todos os momentos de amizade e descontração. Pelo apoio e motivação que sempre me dispensou.

Ao Dr. Domingos Alves, pela receptividade no início de nossa jornada em São Carlos. Agradeço também pelas discussões entusiasmadas sobre os temas relacionados à modelagem biológica, que com certeza me ajudaram muito no aprendizado.

Aos amigos e ex-amigos de grupo: Camilo Neto; Daniela Barbatto; Fernando Fagundes (BIOS Gates); Milton Sonoda; Alexandre Colato; Cláudia Pio; Fernando Crepaldi; Leonardo Maia; Ana Tereza e Daniela Botelho.

Aos amigos do IFSC: Eduardo Ribeiro Azevedo; Vivaldo Leira Campo; João Vítor Batista Ferreira; Luís Roberto Ramos e Marcelo Ferreira da Silva.

À Maria Cristina (secretária de grupo) e Wladerez (secretária da pós-graduação) que sempre se mostraram prestativas e atenciosas.

Aos amigos que temos em São Carlos fora do meio acadêmico e que sempre nos ajudaram muito a tornar São Carlos uma cidade mais agradável para nossa convivência.

Por fim gostaria de agradecer à **Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)** pelo apoio financeiro durante todo esse projeto de doutorado.

Resumo

O paradigma de sistemas de moléculas auto-replicas é o modelo de quase-espécies, no qual as moléculas são representadas por seqüências binárias de tamanho L e o mecanismo de replicação é suposto imperfeito. Em particular, cada seqüência é gerada corretamente com probabilidade $Q = q^L$, onde q é a probabilidade de cópia exata por dígito. Um dos resultados mais intrigantes no modelo para o relevo de replicação de pico único, no qual há apenas um tipo de molécula com vantagem seletiva a em relação aos outros tipos, é a observação de um limiar de erro a partir do qual toda informação biológica relevante é perdida. A transição de limiar de erro verificada para $Q_c = 1/a$ pode ser visualizada como uma transição de fase do tipo ordem-desordem. Verificamos que a largura dessa transição decresce com L de acordo com L^{-1} . Concluimos também que as grandezas físicas de interesse são bem descritas por meio de funções de escala. Elaboramos ainda uma versão estocástica (isto é, tamanho de população N finito) para o modelo de quase-espécies, no qual a dinâmica é descrita por uma cadeia de Markov. Mostramos que o tempo característico τ para o desaparecimento de seqüências mestras na população obedece uma relação de escala bem definida. A transição em nosso modelo é constatada através da divergência de τ em Q_c no limite de $N \rightarrow \infty$, sendo que a largura da transição decresce de acordo com $N^{-1/2}$. Em nossa abordagem não utilizamos nenhuma definição arbitrária para o limiar de erro para população finita.

Como solução para o problema da crise de informação associada ao limiar de erro estudamos o modelo de hiperciclos. Neste modelo, as macromoléculas se replicam com o auxílio de outros membros do hiperciclo por meio do mecanismo de catálise. Estudamos analiticamente a propagação de erro no hiperciclo e obtemos os diagramas de fases no espaço de parâmetros para vários tamanhos de hiperciclo n . Esses diagramas descrevem as regiões de estabilidade das diversas soluções de estado estacionário do sistema. Constatamos que para hiperciclos com $n \leq 4$ existe um limiar de erro menor que aquele verificado no modelo de quase-espécies. Desde que o suporte catalítico realizado por uma molécula no hiperciclo pode ser considerado de fato um comportamento altruísta, modelos para evolução do altruísmo como, a teoria de seleção de grupos, têm sido utilizados no contexto de evolução pré-biótica. Aqui investigamos a evolução da produção de enzimas e os efeitos de sinergia utilizando esses conceitos.

Abstract

The quasispecies model is the paradigm of systems composed of self-replicating molecules which are represented by sequences of fixed length L . The replication machinery is assumed to be imperfect. Particularly, each sequence is copied exactly with probability $Q = q^L$, where q denotes the probability of exact copy per digit. One of the most intriguing results of the model for the single-peak replication landscape, which considers the existence of a master sequence that has a selective advantage a in comparison to the other types, is the occurrence of an error threshold phenomenon beyond the biological information is completely lost. The error threshold transition can be viewed as an order-disorder phase-transition. We investigated the sharpness of the threshold and found that its characteristics persist across a range of Q of order L^{-1} about Q_c . Other physical quantities of interest are also well described by universal functions. We formulate a stochastic version (i.e., finite population size N) for the quasispecies model, in which the dynamics is defined by a Markov chain. We show that the characteristic time τ for the disappearance of master sequences in the population obeys a well defined scaling relation. The transition in this model is signaled by the divergence of τ at Q_c in the limit $N \rightarrow \infty$, and the sharpness of the transition decreases like $N^{-1/2}$. In our approach we do not use any arbitrary definition of the error threshold for finite population.

As a solution for the information crisis associated to the error threshold, here we considered the hypercycle model, where the macromolecules self-replicate with the catalytic support of the other members of the hypercycle. We study analytically the error propagation in the hypercycle and obtain the phases diagrams in the space of parameters for several hypercycle sizes n . These diagrams describe the stability regions of the steady state solutions of the system. We find that for hypercycle size $n \leq 4$ the error threshold is smaller than that for the quasispecies model. Since the catalytic support developed by a molecule is in fact an altruistic behavior, some models for the evolution of altruism, for instance, the selection group theory, have been investigated in the context of pre-biotic evolution. Specifically, here we analyse the evolution of enzyme production and the effects of synergism.

Conteúdo

1	Introdução	15
2	O modelo de quase-espécies	22
2.1	Descrição do modelo	23
2.1.1	Função seletiva de pico único	25
2.2	Estudo de escala de tamanho finito para o modelo de quase-espécies	30
3	Uma abordagem para população finita	38
3.1	Descrição do modelo	39
3.2	Dinâmica do modelo	40
3.3	A matriz de transição T	43
3.4	Tempo de relaxação	45
3.5	Cálculo dos momentos	47
3.6	Resultados	49
3.6.1	O limite $a \rightarrow \infty$	52
3.7	Distribuição semi-gaussiana para o número de seqüências mestras	55
3.8	Tempo de extinção	62
4	Hiperciclos	66
4.1	O modelo	69
4.2	O espaço de concentrações <i>Simplex</i>	72
4.3	Análise de alguns sistemas concretos	75
4.3.1	Competidores Independentes	75
4.3.2	Cadeias catalíticas	77
4.3.3	Ciclos catalíticos	80

4.4	Estudo de Propagação de Erro no Hiper ciclo	83
4.4.1	O modelo	83
4.4.2	Análise de pontos fixos	85
4.4.3	Análise de estabilidade	89
4.4.4	Discussão	90
5	Compartimentalização em Evolução Pré-biótica	96
5.1	O Modelo	98
5.2	Evolução de produção de enzimas	100
5.2.1	Análise do modelo - Determinação de pontos fixos e cálculo de es- tabilidade	102
5.2.2	Tamanho de demes variável	106
5.3	Sinergia	108
5.3.1	Análise do modelo - Determinação de pontos fixos e cálculo de es- tabilidade	111
5.3.2	Tamanho de demes variável	114
6	Evolução em Relevos Rugosos	119
6.1	Redes neutras e competição entre diferentes quase-espécies	120
6.1.1	Relevo de dois picos com neutralidade	123
6.2	O modelo NK	129
6.2.1	Descrição do modelo	129
6.2.2	Catástrofe de erro	131
6.2.3	Velocidade de Adaptação	134
7	Conclusão	136
	Apêndice	142
	Apêndice	142
A	Solução Analítica das Eqs. do modelo de quase-espécies	142
	Apêndice	145
B	Cálculo dos elementos da matriz M de mutação entre classes	146

Apêndice	147
C Método das potências ou iteração vetorial	148
C.1 O autovalor dominante e o autovetor associado de uma matriz	148
C.1.1 Cálculo do segundo autovalor	149
Apêndice	151
D Análise de estabilidade local de pontos fixos	152
Bibliografia	156

Lista de Figuras

2.1	Distribuição da concentração de classes de seqüências como função da probabilidade de erro $(1-q)$ por dígito. Neste exemplo, temos $L = 50$ e $a = 10$.	27
2.2	Distância de Hamming normalizada entre a seqüência mestra e a população, d , como função da probabilidade normalizada de replicação exata Q/Q_c para $a = 10$, e $L = 70$ ($\square$), 100 ($\bigcirc$), 120 ($\triangle$) e 150 ($\times$).	33
2.3	Desvio quadrático médio σ^2 como função da probabilidade normalizada de replicação exata. A figura no interior ilustra o procedimento utilizado para estimar a razão γ/ν . Os parâmetros e convenções são os mesmos da Figura 2.2.	34
2.4	Distância de Hamming normalizada como função da probabilidade de replicação exata reescalada. Os parâmetros são $1/\nu = 1$ e (de cima para baixo) $a = 10, 20$ e 50 . A convenção é a mesma utilizada na Figura 2.2.	35
2.5	Desvio quadrático médio reescalado como função da probabilidade de replicação exata reescalada. Os parâmetros são $1/\nu = 1$, $\gamma/\nu = 1.96$, e (de cima para baixo na região do pico) $a = 10, 20$ e 50 . A convenção é a mesma utilizada na Figura 2.2.	36
2.6	Colapso das funções dependentes de a ilustradas na Fig. 2.4. A convenção e os parâmetros são os mesmos para a Fig. 2.2.	37
3.1	A área azul representa a fração de seqüências mestras na população ponderada pelo seu valor adaptativo a , e a área vermelha a fração da população composta pelas demais classes.	42
3.2	Comportamento do tempo de relaxação τ como função de Q para vários valores de N . Neste gráfico o valor seletivo da classe de mestras é $a = 2$. A curva sólida representa o resultado analítico dado pela Eq. (3.32), isto é, $\tau = -\frac{1}{\ln(2Q)}$.	50

3.3	Tempo de relaxação τ versus Q para vários valores de N . Neste gráfico o valor seletivo da classe de mestras é $a = 20$. A curva sólida representa o resultado analítico dado pela Eq. (3.32), isto é, $\tau = -\frac{1}{\ln(20Q)}$	51
3.4	Logaritmo do tempo de relaxação, $\ln \tau$, calculado em $Q = Q_c = 1/a$ como função do logaritmo do tamanho da população, $\ln N$, para valores de $a = 2(\bigcirc)$, $a = 10(\triangle)$ e $a = 50(\square)$	52
3.5	Reescala do logaritmo do tempo de relaxação $\ln \tau / \ln N^{1/2}$ como função de $(Q - Q_c)N^{1/2}$ para valores de (de cima para baixo em $Q = Q_c$) $a = 2, 10$ e 50 . A convenção adotada é $N = 200(\diamond), 300(\square), 400(\bigcirc), 500(\triangle)$, e $600(\times)$. Para $a = 10$ e 50 apenas os dados para $N \geq 400$ estão presentes.	53
3.6	Evolução exata da concentração de seqüências mestras no tempo. Os parâmetros utilizados são $a = 2$ e $N = 300$	56
3.7	δ como função de Q para alguns valores de N . Neste gráfico temos $a = 10$	57
3.8	σ^2 como função de Q para alguns valores de N . Neste gráfico temos $a = 10$	58
3.9	Concentração média do número de mestras $\langle y(t) \rangle$ versus <i>tempo</i> para alguns valores de N . Como parâmetros temos $a = 10$ e $Q = 0.15$. A transição para $N \rightarrow \infty$ ocorre em $Q_c = 0.1$	58
3.10	Razão δ/σ^2 para valor seletivo $a = 10$ e alguns valores de N . A linha sólida representa o valor teórico aproximado.	60
3.11	σ^2 versus Q para alguns valores de N . Nesta Figura temos $a = 10$. As linhas sólidas representam os valores analíticos de acordo com a distribuição semi-Gaussiana para cada valor de N	61
3.12	δ versus Q para alguns valores de N . Nesta Figura temos $a = 10$. As linhas sólidas representam os valores analíticos de acordo com a distribuição semi-gaussiana para cada valor de N	62
3.13	Distribuição de probabilidade de evento de extinção das seqüências mestras P_{ext} versus t	64
3.14	Tempo de extinção τ_{ext} versus Q para alguns valores de N . A linha sólida representa os valores numéricos exatos. Os pontos são os dados obtidos por meio de simulação.	65

4.1	O hiperciclo. Cada uma das unidades A, B, C e D é um replicador. A taxa de replicação de cada unidade é uma função crescente da concentração da unidade imediatamente anterior. Assim a taxa de replicação de B é uma função da concentração de A, e assim por diante.	69
4.2	Um simplex para 3 espécies.	73
4.3	O simplex 3-espécies inserido no espaço tri-dimensional.	74
4.4	Localização do ponto $[0.25, 0.25, 0.5]$ em um simplex 3-espécies.	75
4.5	Cadeia catalítica de tamanho n	79
4.6	Um hiperciclo de 4 espécies.	81
4.7	Um hiperciclo de 6 espécies.	82
4.8	Sistema composto de um hiperciclo de tamanho $n = 3$ e sua cauda de erro I_e . As flechas finas representam as reações de auto-replicação não catalizadas e as flechas grossas indicam a auto-replicação cataliticamente assistida pela seqüência vizinha.	85
4.9	Diagrama de fases no espaço (K, Q) para $n = 4$ e $a = 10$ exibindo as regiões de estabilidade dos diversos estados de coexistência. Os números entre parênteses indicam o números de espécies coexistentes. Regiões de biestabilidade aparecem para $K \geq K_h$. As linhas finas são (de baixo para cima) Q_1, Q_2, Q_3 e Q_4 . A linha cheia é Q_h	92
4.10	O mesmo que Fig. 4.9, porém $n = 5$. Não há pontos fixos estáveis acima da linha tracejada. As curvas sólidas são (de baixo para cima) Q_1, Q_2, Q_3, Q_4 e Q_5	93
4.11	Evolução temporal das cinco concentrações dos elementos compondo o hiperciclo de tamanho $n = 5$ para $a = 10$, $Q = 1$, e (a) $K = 37$ (dentro da região de estabilidade) e (b) $K = 40$ (fora da região de estabilidade). O estado inicial é $x_i = 0.2 \forall i$	93
4.12	Evolução temporal das cinco concentrações dos elementos compondo uma cadeia de tamanho $n = 5$. Os parâmetros e o estado inicial inicial são os mesmos utilizados para 4.11(b).	95

5.1	Frequência de estado estacionário de replicadores tipo A para $u = 0$, $N = 5$ e (da esquerda para direita) $k_B/r = 0, 2, 4, 6, 8$ e 10 . As frequências iniciais são (a) $p_0 = 0.999$ e (b) $p_0 = 0.001$. As três primeiras linhas na parte (a) colapsam em uma única curva na parte (b).	104
5.2	Diagramas de fases para $N = 5$ e $r = 0.1$ exibindo as regiões de estabilidade dos diferentes pontos fixos para (a) $u = 0$, (b) $u = 0.005$, (c) $u = 0.01$ e (d) $u = 0.0158$. O ponto de interseção toca o eixo coordenada no ponto $k_A = 5/3$	105
5.3	Frequência de estado estacionário de replicadores tipo A na população para $u = 0.005$, $N = 5$, $r = 0.1$, e (da esquerda para direita) $k_B/r = 0, 2, 2.9, 6, 8$ e 10 . As frequências iniciais são (a) $p_0 = 1$ e (b) $p_0 = 0$	106
5.4	Coordenadas do ponto crítico (a) k_B^c/r e (b) k_A^c/r como funções da taxa de mutação u para $N = 5$ e (da esquerda para direita) $r = 0.1, 0.2, 0.3, 0.5, 0.8$, e 1.0 . Nos pontos onde $k_B^c/r = 0$ encontramos $k_A^c/r = 5/3$	107
5.5	Taxa de mutação u_e além do qual as transições descontínuas desaparecem como função do custo altruísta r para (de baixo para cima) $N = 5, 10, 20, 50$ e ∞	108
5.6	Estado estacionário p^* da equação (5.27) em função de c , para alguns valores de r , e valores fixos de $N = 20$ e $i_m = 1$	112
5.7	Linhas de transição para $N = 20$, $u = 0$ e, de cima para baixo, $i_m = 1, 2, 4, 10$ (linhas sólidas), e $i_m = 19, 17$ (linhas tracejadas). As curvas para $i_m = 20$ e $i_m = 11$ coincidem com aqueles obtidos para $i_m = 1$ e $i_m = 10$, respectivamente.	114
5.8	Diagrama de fases para $N = 20$ e $i_m = 5$ exibindo as regiões de estabilidade dos diferentes pontos fixos para (a) $u = 0$, (b) $u = 0.0085$, (c) $u = 0.0282$ e (d) $u = 0.0330$	115
5.9	Taxa de mutação u_e , além do qual as transições descontínuas desaparecem, como função de i_m para $N = 5$ ($\triangle$), $N = 10$ (∇), $N = 20$ ($\square$) e $N = 30$ ($\diamond$).	116
6.1	Hipercubos representando o espaço de seqüências para seqüências binárias de comprimento de 1 a 4.	120
6.2	Enovelamento de uma seqüência de RNA em sua estrutura espacial.	121

6.3	Um esboço do mapeamento de seqüências de RNA em suas estruturas secundárias.	122
6.4	Um esboço do relevo de adaptação em organismos digitais.	123
6.5	Relevo de adaptação de dois picos. Aqui utilizamos seqüências de tamanho $L = 20$. O primeiro pico tem valor seletivo 7 com largura igual 3, ou seja consiste de todas as seqüências com distância de Hamming menor que 3 em relação à seqüência 000...000. Já as seqüências no segundo pico têm valor seletivo 4 e consiste de todas as seqüências com distância de Hamming menor que 5 em relação à seqüência 111...111.	125
6.6	(a) valor seletivo médio, (b) valor seletivo máximo, (c) concentração de seqüências nos picos 1 e 2 do relevo, e por fim (d) concentração de seqüências mortas; todos como função do parâmetro $(1 - q)$	125
6.7	Entropia da população como função do parâmetro $(1 - q)$	126
6.8	Resultados da simulação para organismos digitais. Temos na parte (a) Valor seletivo médio e (b) Valor seletivo máximo da população com função da taxa de mutação.	127
6.9	Resultados da simulação para organismos digitais. Temos na parte a) Entropia e (b) exponencial da entropia.	128
6.10	Cálculo do valor adaptativo F no modelo NK para o caso $N = 4$ e $K = 2$. $N = 4$ genes compõem o genoma cada qual com um alelo A_i que pode assumir os valores 0 ou 1. Para cada um dos 4 alelos um subgenoma é construído tomando o próprio gene e $K = 2$ outros genes, aleatoriamente escolhidos entre os 3 genes restantes no genoma. A cada subgenoma é atribuído um valor f_i tomado aleatoriamente do intervalo $(0, 1]$. O valor seletivo total F para o genoma é então igual a média dos $N = 4$ valores dos subgenomas.	131
6.11	Histograma da distribuição da distância de Hamming entre as seqüências e a seqüência consenso da população. Para esta simulação temos população de tamanho $N = 5000$, seqüências de tamanho $L = 32$, e parâmetro $K = 8$. A probabilidade de uma mutação pontual é igual a: (a) $u = 0.001$, (b) $u = 0.005$, (c) $u = 0.01$ e (d) $u = 0.1$	132

- 6.12 Evolução do valor adaptativo médio da população. Os parâmetros são os mesmos daqueles utilizados na Figura 6.11. A probabilidade de uma mutação pontual é igual a: (a) $u = 0.001$, (b) $u = 0.005$, (c) $u = 0.01$ e (d) $u = 0.1$ 133
- 6.13 Evolução temporal do valor adaptativo médio da população. Consideramos seqüências de tamanho $L = 64$. Os valores de K utilizados na simulação são: (a) $K = 2$, (d) $K = 4$, (c) $K = 8$ e (d) $K = 16$. As cores denotam os diferentes valores da probabilidade de mutação por dígito, e obedecem a seguinte correspondência: $u = 0.00001$ (vermelho), $u = 0.0001$ (azul), $u = 0.001$ (verde), $u = 0.01$ (marrom escuro), $u = 0.1$ (violeta). 135

Capítulo 1

Introdução

O modelo de quase-espécies, proposto por Manfred Eigen [1] há mais de vinte anos atrás, ainda é hoje o paradigma de sistemas de moléculas auto-replicas que competem em um ambiente limitado energeticamente. Eigen foi um dos primeiros a dar uma descrição matemática exata do fluxo de informação em um sistema sob a ação de seleção natural e mutação. O modelo supõe a existência de macromoléculas que têm a capacidade de auto-replicar-se sem o auxílio de mecanismos como a catálise. O modelo de quase-espécies é elaborado com base nas considerações da teoria Darwiniana de seleção natural, porém é totalmente descrito em termos da cinética química de reações, o que permite uma interpretação mais precisa do conceito de organismo mais apto. O modelo não apenas permite um entendimento das limitações físicas do processo de adaptação, como também fornece uma idéia mais clara do papel da estocasticidade neste processo.

No modelo de quase-espécies, a população é suposta infinita e as frequências dos diferentes tipos de moléculas são descritas por equações diferenciais determinísticas. Cada macromolécula é representada por uma sequência binária de tamanho ou comprimento L , ou seja, $\vec{S} = (s_1, s_2, \dots, s_L)$ com $s_\alpha = 0, 1$. A dinâmica do modelo é descrita pelo seguinte conjunto de equações:

$$\frac{dx_i}{dt} = \sum_j W_{ij} x_j - [D_i + \Phi(t)] x_i \quad (1.1)$$

onde D_i denota a taxa de degradação ou morte de moléculas de tipo $i = 1, \dots, 2^L$. Os

elementos $W_{ij} = A_j q^{L-d(i,j)} (1-q)^{d(i,j)}$ definem a taxa com que um elemento tipo j dá origem a um elemento de tipo i , por meio de erros na replicação (mutação). Naturalmente, os elementos diagonais $W_{ii} = A_i q^L$ definem a taxa com que o elemento i reproduz sua cópia perfeita. Os valores desses elementos dependem da taxa de replicação das espécies, A_j , da probabilidade de cópia exata por dígito da seqüência, q ; como também da distância de Hamming $d(i, j)$ entre as seqüências de tipo i e de tipo j . Por distância de Hamming definimos o número de dígitos, ou posições, que diferem entre duas seqüências.

Um dos resultados mais interessantes desse modelo é a verificação da chamada catástrofe de erro, que tem conseqüências extremamente importantes para as teorias da origem da vida, uma vez que impõe um tamanho máximo, ou complexidade máxima, para as moléculas existentes na Terra nas condições pré-bióticas. As estimativas são de que o comprimento máximo dessas moléculas seria de 10 a 10^2 nucleotídeos. Moléculas maiores só poderiam existir com a presença de enzimas específicas para auxiliar na sua replicação, mas tais enzimas só poderiam ser codificadas por moléculas com 10^3 a 10^4 nucleotídeos. Este problema, denominado de paradoxo de Eigen, pode ser posto na seguinte forma: não existem grandes genomas sem enzimas e não existem enzimas sem grandes genomas. As propostas de resolução desta questão fascinante têm ocupado lugar de destaque na agenda da Biologia Evolucionária.

O modelo de quase-espécies apresenta dois regimes distintos: o regime de quase-espécie e o regime estocástico. No primeiro, um aglomerado de mutantes coexistem de forma estável com a seqüência mestra, a saber, a seqüência de maior valor replicativo. Esse regime pode ser visualizado como um regime ordenado. No segundo regime, a informação genética é totalmente perdida e as seqüências ocupam todo o espaço de genótipos com igual probabilidade. Esse é um regime desordenado. A transição entre esses dois regimes pode ser visualizada como uma transição de fase do tipo ordem-desordem verificada em sistemas de redes de spins, estudados extensivamente em mecânica estatística.

No capítulo 2, descrevemos em detalhe o modelo de quase-espécies e exibimos alguns resultados conhecidos na literatura, entre eles o fenômeno da catástrofe de erro. Nesse capítulo, nosso principal objetivo é a completa caracterização da transição de limiar de erro para valores finitos de tamanho de seqüência. Em particular, iremos realizar um estudo de escala de tamanho finito, técnica comumente empregada em mecânica estatística como ferramenta para melhor compreensão das grandezas críticas do sistema. Esta técnica tem sido utilizada em sistemas físicos exibindo transições de fase, principalmente transições de segunda ordem. Vamos determinar os expoentes críticos descrevendo a largura da região onde as características de transição persistem, como também os demais expoentes associados às grandezas físicas de interesse. Um aspecto interessante nesse estudo é que, diferentemente dos sistemas físicos estudados em mecânica estatística, aqui não existe uma dimensão espacial, de forma a associarmos à transição de fase a um ordenamento espacial ou divergência do comprimento de correlação do sistema. Esse mesmo tipo de fenômeno ocorre, por exemplo, em problemas combinatoriais [2], onde mais uma vez não existe um critério geométrico para definição de grandezas análogas ao comprimento de correlação.

Complementando a análise determinística realizada no capítulo 2, vamos estudar uma versão estocástica do modelo de quase-espécies para o relevo de pico único no capítulo 3. Diversos estudos na literatura mostram que quando lidamos com um número finito de indivíduos que se reproduzem assexuadamente, invariavelmente haverá uma redução aleatória progressiva dos indivíduos mais adaptados, um processo denominado de catraca de Müller [3]. Em outras palavras, ocorre um acúmulo contínuo de mutações deletéreas levando a uma redução do valor seletivo médio da população. Neste capítulo, calculamos o tempo característico τ para o desaparecimento de seqüências mestras na população. Utilizando a técnica de escala de tamanho finito, empregada anteriormente, encontramos uma função de escala que incorpora a dependência de τ com o tamanho da população N .

A grande vantagem dessa abordagem é a possibilidade de estudar as propriedades críticas do problema sem uma definição arbitrária do termo limiar de erro para população finita.

O paradoxo de Eigen não é consequência apenas da catástrofe de erro mencionada acima, mas também da impossibilidade de coexistência de moléculas que não sejam muito semelhantes no modelo de quase-espécies. Se tal existência fosse possível, então poderíamos pensar no sistema composto por essas várias espécies moleculares estáveis como a unidade portadora da informação genética. Surge daí a noção de integração de informação que integra o título dessa tese. A coexistência entre elementos bastante distintos pode ser forçada impondo-se um mecanismo de cooperação ou catálise entre as moléculas: Neste caso, a eficiência de replicação de uma dada molécula depende de uma segunda molécula, cuja replicação depende de uma terceira molécula, cuja replicação, por sua vez, depende da presença da primeira molécula, formando assim um ciclo, denominado hiperciclo. Os hiperciclos têm sido postulados como um evento único em evolução pré-biótica e nos dão uma boa explicação para a emergência do código genético, assim como para o aumento do tamanho do genoma. Seu ponto de partida é a observação de que cadeias e ciclos catalíticos são comuns em todos os organismos modernos e, portanto, devem ter evoluído de algum estágio inicial da origem da vida. Embora tenha sido introduzido há mais de vinte anos [4, 5], o modelo de hiperciclos ganhou uma maior plausibilidade com a verificação experimental de atividades catalíticas desenvolvidas por RNAs [6, 7].

Embora, o fenômeno da catástrofe de erro tenha sido considerado a maior motivação para a introdução do modelo de hiperciclos, a grande maioria dos trabalhos encontrados na literatura lidam apenas com os aspectos referentes à coexistência dos elementos do hiperciclo, desde que assumem replicação perfeita das macromoléculas [8–10]. Neste caso, um número arbitrário n de espécies coexistem permanentemente em um estado de equilíbrio dinâmico. Entretanto, se $n > 4$ as concentrações desses diferentes tipos de

elementos variam com o tempo, decrescendo periodicamente para valores bastante reduzidos. Portanto, na prática grandes hiperciclos são extremamente suscetíveis à extinção por meio de flutuações ou perturbações no sistema.

O efeito de replicações errôneas na organização hipercíclica foi investigado introduzindo-se um campo mutação como uma perturbação das equações cinéticas sem mutações [11]. Entretanto, este procedimento não é apropriado para o estudo do fenômeno de limiar de erro, desde que os resultados obtidos não podem ser facilmente comparados com aqueles do modelo de quase-espécies. Dessa forma, uma suposição mais adequada para esse estudo é admitir a existência de uma classe especial de seqüências que não desenvolvem qualquer atividade catalítica, a denominada cauda de erro, e que surgem como consequência da replicação errônea dos elementos do hiperciclo [12–14].

No capítulo 4 estudamos o modelo de hiperciclos. Apresentamos também exemplos de alguns sistemas concretos já estudados na literatura e os seus resultados correspondentes. Nesta parte da tese, temos por finalidade a análise da propagação de erro em um sistema hipercíclico e sua cauda de erro. O sistema é descrito por um conjunto de equações diferenciais que descrevem de forma determinística a evolução do sistema. Investigamos hiperciclos de vários tamanhos, calculando os pontos fixos, que são as soluções de estado estacionário do sistema de equações como também a estabilidade desses pontos. De posse desses resultados, determinamos o diagrama de fases, que mostra as regiões de existência e estabilidade das soluções no espaço de parâmetros.

Uma importante objeção à teoria de hiperciclos foi levantada por Maynard Smith [15, 16]. Essa objeção deve-se ao fato dos hiperciclos serem vulneráveis à invasão de parasitas, isto é, moléculas que se beneficiam do suporte catalítico de elementos do hiperciclo, mas por sua vez não auxiliam na replicação dos outros membros da rede catalítica. Assim, diversas soluções alternativas têm sido discutidas para o modelo de hiperciclos [17–21].

Todas elas supõem que seqüências auto-replicas competem dentro de pequenos compartimentos ou grupos, denominados *demes*, e seus valores seletivos dependem da composição interna do compartimento. Nesse sentido, esses modelos são denominados genericamente de modelos de seleção de grupo. Esses estudos têm por finalidade a compreensão de mecanismos capazes de manter a coexistência de elementos estruturalmente distintos numa população de macromoléculas. Em geral, os modelos de seleção de grupo assumem a existência de elementos que conferem alguma vantagem seletiva ao grupo ao qual pertencem, entretanto isso acarreta um custo à sua taxa de auto-replicação, o que determina um comportamento altruísta. Em resumo, definimos por comportamento altruísta como aquele que é de alguma forma prejudicial ao próprio elemento que o desempenha, mas que proporciona uma vantagem ao grupo do qual é parte. Sob certas condições esse tipo de mecanismo pode garantir a coexistência estável dos elementos envolvidos.

Em formulações mais modernas de seleção de grupo, as *demes* são visualizadas como grupos característicos, nos quais as interações entre os indivíduos efetivamente ocorrem, embora seja permitido aos indivíduos competirem pelos recursos totais disponíveis no meio [22]. Alguns sistemas biológicos reais podem ser descritos adequadamente pelos grupos característicos. Como exemplo, citamos a dinâmica de seleção viral, onde as células desempenham o papel das *demes* acima descritas. No contexto de evolução pré-biótica, a motivação para o advento de tais modelos baseados no encapsulamento de macromoléculas vem do provável cenário para a sopa primitiva, onde certamente a distribuição dessas macromoléculas não era homogênea. De fato, fendas em rochas e partículas de poeira suspensas [23] seriam os habitats locais, dentro dos quais essas moléculas seriam distribuídas. Em particular, Woese [24] e Towe [25] sugeriram um cenário atmosférico para a origem da vida em gotas de água suspensas, que seriam dispersadas por meio da ação do vento.

No capítulo 5, investigamos alguns modelos de compartimentalização em evolução

pré-biótica no contexto da teoria de seleção de grupo. Em nosso modelo, a população é composta de um número infinito de *demes*, cada uma constituída de N indivíduos haplóides que se reproduzem assexuadamente. Como na formulação de Wilson [22], aqui consideramos que são permitidos aos indivíduos competirem e acessarem os recursos totais disponíveis no meio. Entretanto, diferentemente da aproximação de campo médio utilizada, aqui consideramos que o valor seletivo de um indivíduo depende da fração de altruístas que ele experimenta dentro da própria deme. Esse formalismo é então aplicado ao estudo do modelo de evolução para a evolução da produção de enzimas proposto por Michod [17] e também na investigação dos efeitos de sinergia [18].

Apresentamos uma discussão preliminar sobre as conseqüências de considerarmos aspectos reais de relevos de adaptação, como rugosidade e neutralidade. Apresentamos alguns resultados experimentais, obtidos por meio de simulação em plataformas artificiais de evolução. Queremos mostrar que o fenômeno da catástrofe de erro estudado no relevo de pico único também é verificado em relevos rugosos. Para isso, fazemos uma breve discussão utilizando como modelo, o modelo NK de adaptação introduzido por S. Kauffman [26]. Por fim, investigamos como a mutação pode influenciar o processo de otimização evolucionária, acelerando ou não a busca por mínimos locais em relevos rugosos.

Finalmente, no capítulo 7, vamos apresentar as conclusões e perspectivas dos temas abordados em nosso trabalho.

Capítulo 2

O modelo de quase-espécies

Os organismos existentes no nosso planeta são o resultado de bilhões de anos de aperfeiçoamento guiados pelo processo de evolução por seleção natural. Naturalmente, não seria razoável esperar que os sistemas primitivos possuísem a complexidade e sofisticação dos processos bioquímicos que guiam o processo reprodutivo em formas de vida atuais. Por esta razão, no estudo de evolução pré-biótica devemos nos entreter apenas com os processos que são centrais na reprodução de tais sistemas, o que já constitui uma tarefa bastante complexa. Um modelo abrangente de origem da vida deve portanto explicar como a maquinária de replicação e produção de proteínas pode surgir espontaneamente de simples precursores.

Descreveremos abaixo um modelo bastante compacto de um sistema bioquímico primitivo originalmente proposto por Manfred Eigen [1] em 1971, o modelo de *quase-espécies*. Eigen foi um dos primeiros a apresentar uma descrição matemática exata do fluxo de informação em um sistema sob a ação de mutação e seleção. O modelo supõe a existência de moléculas que possuem a habilidade de auto-replicação sem a presença de catálise. Essas moléculas podem ser visualizadas como moléculas de DNA (ácido desoxirribonucléico) ou RNA (ácido ribonucléico), que são as responsáveis pela armazenagem e transmissão de informação genética em todos os organismos. Cada espécie molecular tem suas próprias taxas de reprodução e mortalidade. Em adição, o mecanismo de replicação

é suposto imperfeito, de forma que algumas das proles de uma dada molécula não são cópias perfeitas, ou seja, o mecanismo admite mutações nos genes. O modelo enfatiza a interação entre seleção natural e mutações.

O modelo de quase-espécies é construído com base nas considerações da teoria Darwiniana de seleção natural, porém é totalmente descrito em termos da cinética química de reações, o que permite uma visualização mais precisa de organismo *mais apto*. O modelo não apenas dá um entendimento das limitações físicas do processo de adaptação, como também fornece uma idéia mais clara do papel da estocasticidade no processo.

O objetivo do modelo é tentar compreender como mecanismos básicos de evolução, tais como a seleção natural e a mutação, determinam o comportamento de um sistema. Por exemplo, em que condições a espécie mais adaptada irá predominar? De que forma a limitação de energia e matéria pode modificar o comportamento do sistema? Pode o sistema tolerar qualquer valor de taxa de mutação sem que haja perda de informação genética? Essas são algumas questões fundamentais para entender o processo de adaptação de um sistema auto-replicativo.

2.1 Descrição do modelo

Um modelo matemático para processos evolucionários a nível molecular foi proposto por Manfred Eigen em 1971. Ele considerou um sistema de macromoléculas não-interagentes que são capazes de se auto-reproduzirem por um processo de cópia predisposto a erros. Neste modelo a população é suposta infinita e as frequências relativas dos diferentes tipos de moléculas são descritas por equações diferenciais determinísticas. Nesta formulação cada macromolécula é representada por uma seqüência de símbolos de tamanho fixo L , ou seja, $\vec{S} = (s_1, s_2, \dots, s_L)$. Para cada sítio s_α desta seqüência, são permitidos κ diferentes símbolos (nucleotídeos), cada um representando um tipo diferente de monômero utilizado na construção da macromolécula. No caso de uma molécula de DNA, cada po-

sição da seqüência pode ser ocupada por um dos 4 diferentes tipos de bases: A(adenina), G(guanina), C(citosina), T(timina), e portanto, temos $\kappa = 4$. Por simplicidade, utilizaremos apenas seqüências binárias ($s_\alpha = 0, 1$), distinguindo apenas entre as purinas (G, A) e pirimidinas (C, T) como bases das macromoléculas. O número total de seqüências possíveis é portanto igual a $M = 2^L$. Tomando como exemplo o genoma de um vírus como o HIV que é constituído por cerca de 10000 nucleotídeos, o número de diferentes seqüências possíveis é igual a $M = 4^{10000} \sim 10^{6000}$ e constitui o que denominamos de espaço de genótipos.

Vamos agora introduzir a dinâmica do modelo. As concentrações relativas x_i das moléculas do tipo $i = 1, 2, \dots, 2^L$ evoluem no tempo de acordo com as seguintes equações diferenciais [1, 4]

$$\frac{dx_i}{dt} = \sum_j W_{ij} x_j - [D_i + \Phi(t)] x_i, \quad (2.1)$$

onde D_i define a taxa de mortalidade das moléculas do tipo i , e o termo $\Phi(t)$ descreve um fluxo de diluição que mantém o número de moléculas constante, sendo responsável pelo surgimento de uma interação efetiva entre as moléculas. Este fluxo introduz uma não-linearidade na Eq. (2.1) e é determinado pela condição $\sum_i \frac{dx_i}{dt} = 0$. Supondo $D_i = 0$ para todo i e $\sum_i x_i = 1$ temos que

$$\Phi(t) = \sum_{i,j} W_{ij} x_j. \quad (2.2)$$

Na verdade o fluxo de diluição $\Phi(t)$ é exatamente igual ao valor seletivo médio da população $\bar{E}(t)$, dado por

$$\bar{E}(t) = \sum_i A_i x_i = \sum_{ij} W_{ij} x_j. \quad (2.3)$$

Os elementos da matriz replicação W são dados por

$$W_{ij} = \begin{cases} A_i q^L = A_i Q_{ii} & \text{para } i = j \\ A_j q^{L-d(i,j)} (1-q)^{d(i,j)} = A_j Q_{ij} & \text{para } i \neq j \end{cases}, \quad (2.4)$$

onde A_i é a taxa de replicação ou valor seletivo da molécula de tipo i , sendo, no contexto biológico, proporcional ao número de cópias geradas. A definição das taxas de replicação em todo o espaço de genótipos estabelece o que denominamos de função seletiva. Já $d(i, j)$ é a distância de Hamming entre as seqüências i e j e, é dada pelo número de posições (sítios) que diferem entre as duas seqüências. Quanto menor for a distância entre um determinado par de moléculas, maior será sua afinidade biológica. Em nossa definição $0 \leq q \leq 1$ descreve a precisão de replicação exata por dígito da seqüência, que é suposta constante para todos os dígitos.

A precisão do processo de replicação é descrita pela matriz $Q = \langle Q_{ij} \rangle$. Os elementos de Q são adimensionais e Q_{ij} é simplesmente a probabilidade de se obter uma seqüência $\vec{S}_i$ por meio da replicação imperfeita da seqüência $\vec{S}_j$. Portanto, o termo $W_{ij}x_j$ na Eq. (2.1) define a taxa de replicação da molécula do tipo i devido à replicação errônea da molécula do tipo j . Evidentemente, os elementos diagonais Q_{ii} descrevem o fator de qualidade do mecanismo de replicação (probabilidade de replicação exata de uma seqüência). Os elementos não-diagonais $Q_{ij}(i \neq j)$ são portanto as probabilidades de mutações. Conforme esperado, podemos facilmente mostrar que

$$\sum_i Q_{ij} = 1 \quad \forall \quad j, \quad (2.5)$$

indicando simplesmente que no processo de replicação de uma dada molécula, ou ela é sintetizada corretamente ou então erroneamente. A solução analítica do conjunto de Equações (2.1) é discutida no apêndice A.

2.1.1 Função seletiva de pico único

De forma a simplificar drasticamente a descrição e a dimensão do problema, pois o número possível de seqüências binárias $M = 2^L$ é bastante restritivo, devemos formar classes de

seqüências dentro da distribuição de quase-espécies. Essas classes são definidas por meio da distância de Hamming entre a seqüência mestra, a saber $\vec{S}_0 = (0, 0, \dots, 0)$, e a seqüência em consideração. Dessa forma, a classe 0 contém exclusivamente a seqüência mestra; a classe 1 todos os $\binom{L}{1}$ mutantes com $s_\alpha = 1$ em uma única posição da seqüência; a classe 2, todos os $\binom{L}{2}$ mutantes com $s_\alpha = 1$ em apenas duas posições da seqüência, e assim por diante.

Devemos agora, fazer uma suposição de forma a reduzirmos o sistema de 2^L equações diferenciais descrevendo a evolução de seqüências individuais, para um sistema de $(L + 1)$ equações diferenciais descrevendo a evolução das classes de seqüências: todas as taxas de replicação são supostas constantes dentro de uma mesma classe, ou seja,

classe 0, $A_0 = A'_0$;

classe 1, $A_1 = A_2 = \dots = A_L = A'_1$;

classe 2, $A_{L+1} = A_{L+2} = \dots = A_{L+\binom{L}{2}} = A'_2$;

classe k , $A_{\binom{L}{1}+\binom{L}{2}+\dots+\binom{L}{k-1}+1} = \dots = A_{\binom{L}{1}+\binom{L}{2}+\dots+\binom{L}{k}} = A'_k$ e assim por diante.

Feita a suposição acima e estratificando a população em classes que são diferenciadas por meio da distância de Hamming entre as seqüências e a seqüência mestra, as concentrações relativas Y_P das moléculas de classe $P = 0, \dots, L$ evoluem no tempo de acordo com as seguintes equações diferenciais [27]

$$\frac{dY_P}{dt} = \sum_{R=0}^L A'_R M_{PR} Y_R - \Phi(t) Y_P. \quad (2.6)$$

Claramente a conservação da concentração total $\sum_P Y_P = 1$ é assegurada. A matriz $M = \langle M_{PR} \rangle$ agora descreve a probabilidade de mutação de uma molécula de classe R para uma molécula de classe P , e é dada por [27–30]

$$M_{PR} = \sum_{I=I_l}^{I_u} \binom{R}{I} \binom{L-R}{P-I} q^{L-P-R-2I} (1-q)^{P+R-2I}, \quad (2.7)$$

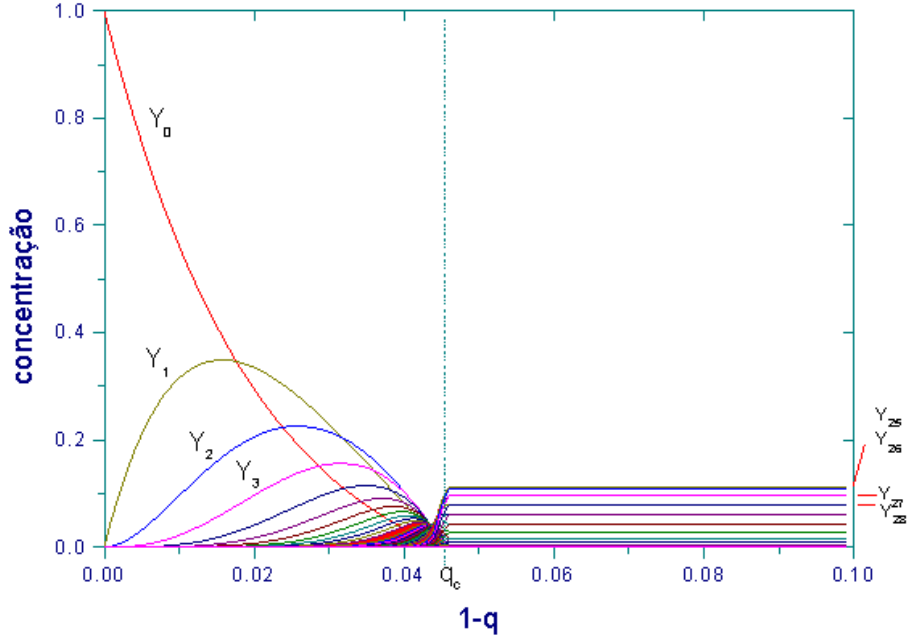


Figura 2.1: Distribuição da concentração de classes de seqüências como função da probabilidade de erro ($1-q$) por dígito. Neste exemplo, temos $L = 50$ e $a = 10$.

onde $I_l = \max(0, P + R - L)$ e $I_u = \min(P, R)$. A derivação detalhada desta equação é apresentada no apêndice B.

Vamos a seguir analisar as principais propriedades e aspectos do modelo de quase-espécies definindo uma das mais simples funções seletivas para este modelo: a função seletiva de pico único. Mais especificamente, consideramos uma distribuição dos valores seletivos para as classes de seqüências atribuindo

$$A'_P = \begin{cases} a & \text{para } P = 0 \\ 1 & \text{para } P \neq 0 \end{cases}, \quad (2.8)$$

ou seja, apenas a seqüência mestra $\vec{S}_0 = (0, 0, \dots, 0)$ possui um valor seletivo diferenciado. Neste contexto o parâmetro a é denominado de vantagem seletiva da seqüência mestra. Na Figura 2.1 exibimos a solução das Eqs. (2.6) no regime estacionário para valores de parâmetros $L = 50$ e $a = 10$. Como podemos verificar, o modelo cinético de replicação e mutação dá origem a um estreito limiar de erro. Neste limiar podemos distinguir uma

transição entre dois regimes distintos na composição da população: no primeiro regime para $q > q_c$, observamos uma distribuição de mutantes em torno de $\vec{S}_0$, aglomerado este composto por classes de seqüências biologicamente mais afins (menor distância de Hamming em relação à mestra). A este regime Eigen denominou de *regime de quase-espécies*. O segundo regime para $q < q_c$, denominado de *regime estocástico*, é caracterizado por uma distribuição uniforme das seqüências, onde os 2^L diferentes tipos de moléculas surgem na mesma proporção, e portanto as concentrações das classes de seqüências são dadas por $Y_P = \frac{1}{2^L} \binom{L}{P}$. Ou seja, a população realiza uma caminhada aleatória por todo o espaço de genótipos. Esse limiar de erro q_c , onde a estrutura da população muda rapidamente de uma distribuição onde a vantagem seletiva parece ter um papel fundamental (regime de quase-espécies) para uma distribuição aleatória das seqüências onde essa informação genética é totalmente perdida, tem sido observado em uma grande variedade de sistemas [31, 32], representados pela diversas estruturas da matriz W . Esse fenômeno é hoje bem aceito como um aspecto intrínseco e pendente do modelo de quase-espécies, sendo denominado de catástrofe de erro.

A transição do limiar de erro já é bastante abrupta para tamanhos de seqüências $L = 50$ conforme pode ser observado na Figura 2.1, tornando-se ainda mais estreita para L maiores. Um enfoque maior e uma análise mais aprofundada dessa transição será exibida na próxima seção, onde determinaremos numericamente a dependência desse estreitamento e de outras grandezas de interesse com o tamanho das seqüências L .

Uma estimativa grosseira para q_c pode ser encontrada por meio da seguinte observação: para que a seqüência mestre $\vec{S}_0$ seja competitiva em relação à outros mutantes, sua taxa de produção deve ser maior que a taxa de produção média de todos os outros mutantes $\vec{A}'_{P \neq 0}$. Ou seja,

$$W_{00} > \overline{A'}_{P \neq 0} = \sum_{P \neq 0} A'_P Y_P, \quad (2.9)$$

e Y_P denota as concentrações de regime estacionário das classes de seqüências. Desde que $W_{00} = A'_0 Q$ e, $Q = q^L$ é a probabilidade de replicação exata, encontramos que esta probabilidade deve satisfazer à condição

$$Q > \frac{\overline{A'}_{P \neq 0}}{A'_0} = \frac{1}{a}. \quad (2.10)$$

De acordo com isto, a transição para o regime estocástico, onde todas as seqüências coexistem na mesma proporção, ocorre para valores próximos a

$$Q_c = \frac{1}{a}. \quad (2.11)$$

Para seqüências infinitas ($L \rightarrow \infty$), essa estimativa é exata pois $\overline{A'}_{P \neq 0} = \sum_{P \neq 0} A'_P Y_P = 1$ e, $Y_0 = \frac{1}{2^L}$ no regime estocástico e portanto pode ser desprezada. Um aspecto interessante do fenômeno da catástrofe de erro pode ser visualizado diretamente da Eq. (2.10), que pode ser reescrita como

$$L \leq -\frac{\ln a}{\ln q}. \quad (2.12)$$

Portanto, de modo a não permitir a perda da informação genética de um determinado sistema, existe uma restrição ao tamanho máximo das seqüências que o compõem. Este aspecto surpreendente no modelo de quase-espécies tem conseqüências extremamente importantes para as teorias de origem da vida, uma vez que impõe uma complexidade máxima para as moléculas existentes na Terra nas condições pré-bióticas. As estimativas são de que o comprimento máximo dessas moléculas seria de 10 a 10^2 nucleotídeos [4, 15]. Moléculas maiores só poderiam existir com a presença de enzimas específicas para auxiliar na sua replicação, mas tais enzimas só poderiam ser codificadas por moléculas com 10^3

a 10^4 nucleotídeos. Esse problema, que na literatura é denominado paradoxo de Eigen, pode ser de certa forma minimizado por meio do conceito de integração de informação. O conceito de integração de informação admite a coexistência de várias espécies moleculares estáveis que não sejam muito semelhantes como a unidade portadora de informação genética. Essa coexistência pode ser forçada impondo-se um mecanismo de cooperação ou catálise entre as moléculas, no qual a eficiência de replicação de uma dada molécula depende da presença de uma segunda molécula, que por sua vez depende de uma terceira molécula, cuja replicação depende da primeira, formando assim um ciclo, denominado *hiperciclo*, que será alvo de estudo em capítulo posterior.

2.2 Estudo de escala de tamanho finito para o modelo de quase-espécies

Nesta seção caracterizaremos a transição de fase exibida no modelo de quase-espécies, já mencionada anteriormente, e que tem sido alvo de diversas análises [33–35]. Neste trabalho utilizamos um procedimento largamente utilizado como poderosa ferramenta no estudo de sistemas que exibem transições de fase, principalmente em sistemas exibindo transições de segunda ordem. Esse procedimento consiste na determinação de funções de escala, isto é, independam do tamanho do sistema. A aplicação dessa análise permitirá um melhor entendimento das grandezas críticas no modelo de quase-espécies. Como exemplo, similarmente à definição de temperatura crítica para sistemas de rede de spins de tamanho finito, não existe uma caracterização conclusiva do termo limiar de erro para seqüências de tamanho finito.

Nesse estudo vamos nos concentrar na função seletiva de pico único, definida anteriormente. Como já enfatizado e discutido anteriormente, dois regimes são observados na composição da população quando variamos o parâmetro q : o regime de quase-espécies e

o regime estocástico (ver Figura 2.1). A transição entre esses dois regimes ocorre no limiar de erro q_c . Para estudarmos esta transição para valores grandes de L é mais conveniente trabalharmos com o parâmetro $Q = q^L$, que define a probabilidade de replicação exata de uma sequência. Para o limite $L \rightarrow \infty$ esta transição ocorre em [1, 36]

$$Q_c = \frac{1}{a}. \quad (2.13)$$

Embora exista um consenso que uma transição de fase ocorre no limite $L \rightarrow \infty$ [33–35], há alguma discordância com relação à ordem desta transição. Estudos baseados no mapeamento do modelo de Eigen para um sistema de Ising bi-dimensional semi-infinito com interações entre primeiros vizinhos em uma única direção conjecturam que o parâmetro de ordem relevante ao problema, a saber, a distância de Hamming normalizada d entre a sequência mestra e o restante da população, exibe uma transição de fase contínua em $Q = Q_c$. Entretanto, devido à enorme dificuldade de resolução das equações auto-consistentes que descrevem as propriedades de equilíbrio de superfície, a análise foi restrita a $L = 20$ [34]. Um outro estudo, um mapeamento alternativo das Eqs. (2.1) para um problema de polímeros direcionados em um meio aleatório indica que a concentração de sequências mestras exibe uma descontinuidade em $Q = Q_c$ [35]. Desde que esse método permite a solução exata do modelo de quase-espécies para a função seletiva de pico único com tamanhos de sequência L genéricos, o resultado implica que a transição para $L \rightarrow \infty$ é definitivamente de primeira ordem.

Nosso objetivo neste trabalho é verificar os efeitos de tamanho finito na transição de fase desse sistema. De particular interesse é a determinação da largura da transição, ou seja, o intervalo de Q próximo a Q_c onde a característica da transição persiste.

Como devemos esperar, a largura dessa região tende a zero de acordo com $L^{-\frac{1}{\nu}}$ quando $L \rightarrow \infty$. Nosso objetivo é estimar o valor do expoente $\nu \geq 0$ usando a técnica de escala de tamanho finito [37]. Nossa abordagem segue a mesma linha de escala de

tamanho finito para problemas combinatoriais [2], onde não há critério geométrico para definição de grandezas análogas ao comprimento de correlação ξ , e desta forma o sucesso do método na determinação da dependência do parâmetro de ordem com alguma escala não pode ser atribuído à divergência do comprimento de correlação espacial.

Ao invés de mapearmos as equações de cinética química em problemas de mecânica estatística de equilíbrio, recorreremos a uma abordagem bem mais simples e direta, que é a solução numérica exata dessas equações no regime estacionário, onde utilizamos comprimentos de seqüência de até $L = 150$.

O procedimento utilizado por nós para a solução numérica das Eqs. (6.1) no regime estacionário $dY_P/dt = 0$ é bastante simples. As concentrações de estado estacionário Y_P ($P = 0, \dots, L$) são encontradas facilmente por meio do seguinte sistema de $(L + 1)$ equações de recorrência [28]

$$Y_P(t + 1) = \frac{\sum_{R=0}^L M_{PR} Y_R(t) + (a - 1) Y_0(t) M_{P0}}{1 + Y_0(t)(a - 1)}, \quad (2.14)$$

desde que o estado estacionário do modelo de quase-espécies apresenta um único ponto fixo estável (ver apêndice A).

As quantidades relevantes na descrição da estrutura da população são a distância de Hamming normalizada entre a seqüência mestra e o restante da população, definida por

$$d = \frac{1}{L} \sum_{P=0}^L P Y_P, \quad (2.15)$$

e o desvio quadrático médio em torno de d ,

$$\sigma^2 = L^2 \sum_{P=0}^L \left(\frac{P}{L} - d \right)^2 Y_P. \quad (2.16)$$

Aqui d e σ^2 são os análogos à magnetização e susceptibilidade para um sistema de rede de spins. Nas Figuras 2.3 e 2.4 exibimos o comportamento dessas duas grandezas como funções da probabilidade de replicação exata normalizada Q/Q_c . Como esperado, os

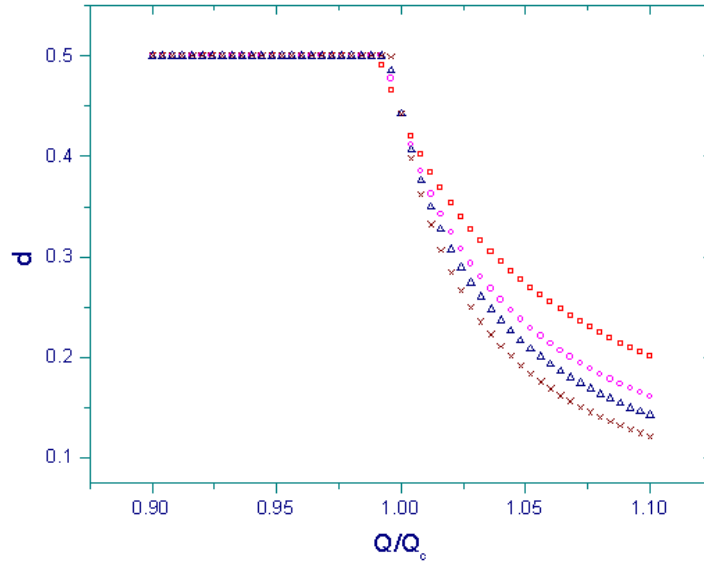


Figura 2.2: Distância de Hamming normalizada entre a seqüência mestra e a população, d , como função da probabilidade normalizada de replicação exata Q/Q_c para $a = 10$, e $L = 70$ ($\square$), 100 ($\circ$), 120 ($\triangle$) e 150 ($\times$).

resultados da Fig. 2.2 mostram o estreitamento da transição com o aumento de L . Além do mais, todas as curvas interceptam-se em um único ponto, o ponto crítico $Q = Q_c$. Este é um resultado um tanto quanto inesperado e mostrou ser bastante útil na localização do limiar de erro, pois em $Q = Q_c$ temos que o parâmetro de ordem independe de L . Isto nos leva a acreditar que a distância de Hamming normalizada d apresenta uma lei de escala do tipo [37, 38]

$$d(L, Q) = \tilde{d}(L/\xi_Q) = \tilde{d}((Q - Q_c)L^{1/\nu}), \quad (2.17)$$

onde ξ_Q é o análogo ao comprimento de correlação em sistemas de rede de spins, e que também apresenta uma divergência em $Q = Q_c$. Propondo uma definição análoga a verificada para problemas onde o critério geométrico está presente, temos que

$$\xi_Q = \left| 1 - \frac{Q}{Q_c} \right|^{-\nu}. \quad (2.18)$$

Queremos salientar que ξ_Q não está associado a nenhuma definição de comprimento no

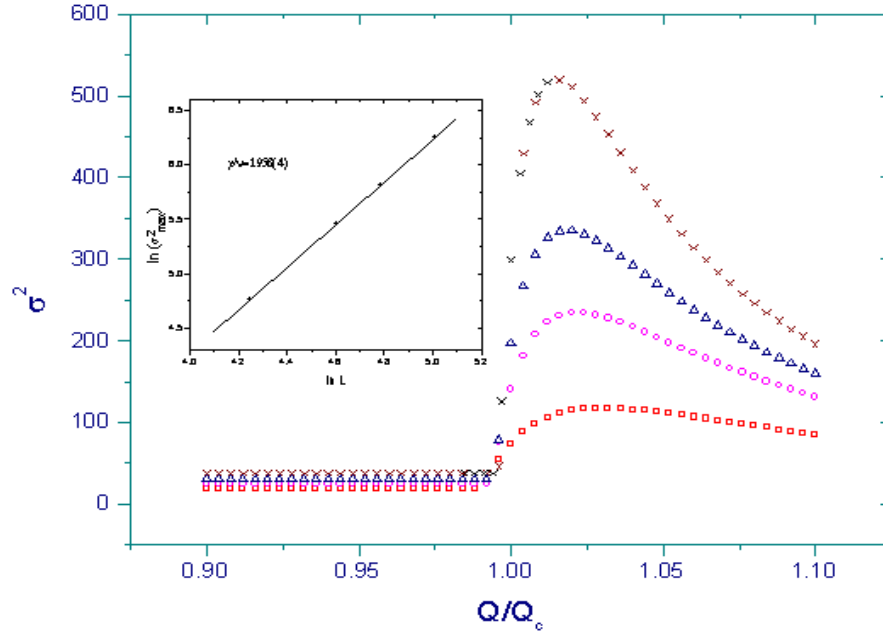


Figura 2.3: Desvio quadrático médio σ^2 como função da probabilidade normalizada de replicação exata. A figura no interior ilustra o procedimento utilizado para estimar a razão γ/ν . Os parâmetros e convenções são os mesmos da Figura 2.2.

sentido físico da palavra. Estes mesmos comportamentos e definições são bastante semelhantes àqueles verificados para o parâmetro de ordem $P_s(p)$ definido para o problema de percolação [37].

As curvas exibidas na Fig. 2.3 indicam que a altura do pico de σ^2 , aqui representada por $\sigma_{\max}^2$, aumenta com L exibindo uma lei de potência da forma $L^{\frac{\gamma}{\nu}}$. A razão dos expoentes $\frac{\gamma}{\nu}$ pode também ser deduzida por meio de um gráfico log-log de $\sigma_{\max}^2$ versus L . Esta verificação nos leva mais uma vez a definição de uma relação de escala para $\sigma^2(Q)$

$$\sigma^2(L, Q) = L^{\frac{\gamma}{\nu}} \tilde{\sigma}^2(z) \quad (2.19)$$

onde

$$z = \varepsilon L^{1/\nu} \quad (2.20)$$

e

$$\varepsilon = (Q - Q_c)/Q_c \quad (2.21)$$

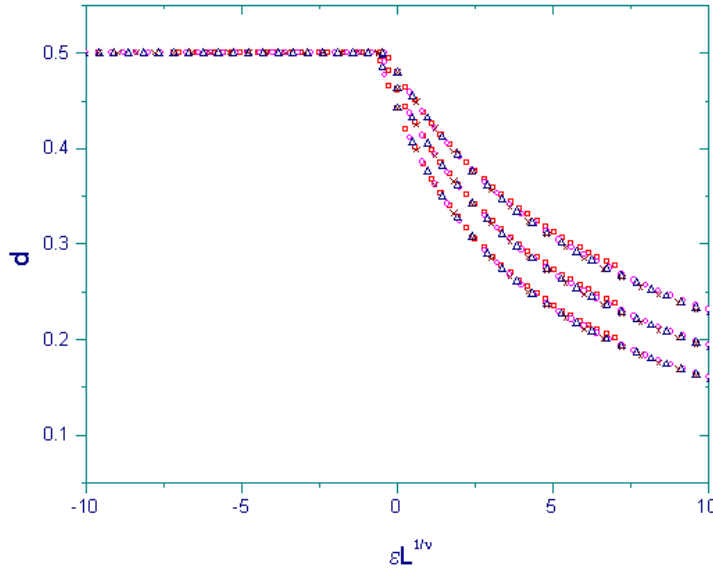


Figura 2.4: Distância de Hamming normalizada como função da probabilidade de replicação exata reescalada. Os parâmetros são $1/\nu = 1$ e (de cima para baixo) $a = 10, 20$ e 50 . A convenção é a mesma utilizada na Figura 2.2.

é a probabilidade reduzida de replicação exata. O resultado $\gamma/\nu = 1.96$ está em ótima concordância com o resultado analítico obtido por Galluccio, no qual a amplitude do desvio médio de uma quase-espécie em torno da sequência mestra ($\sqrt{\sigma^2}$) diverge algebricamente com o expoente igual a -1 quando $Q \rightarrow Q_c$ [35].

O expoente $1/\nu$ é determinado por meio do método do colapso de dados [37], ou seja, a obtenção de funções de escala (conforme ilustrado nas Figs. 2.4 e 2.5). Nestas figuras exibimos $\tilde{d}$ e $\tilde{\sigma}^2(z) = L^{-\frac{\gamma}{\nu}} \sigma^2(Q)$ versus $L^{1/\nu} \varepsilon$, respectivamente. O colapso das curvas para diferentes valores de L foi conseguido adotando os expoentes $1/\nu = 1$ e $\gamma/\nu = 1.96$, independentemente do valor da vantagem seletiva a da sequência mestra, indicando assim uma universalidade desses expoentes. Com estes resultados, concluímos que a largura da transição em torno de $Q_c = 1/a$ é da ordem de L^{-1} .

Como pode ser observado, as funções de escala apresentadas nas Figs. 2.4 e 2.5 parecem exibir uma dependência com o parâmetro a na região crítica. Esta dependência

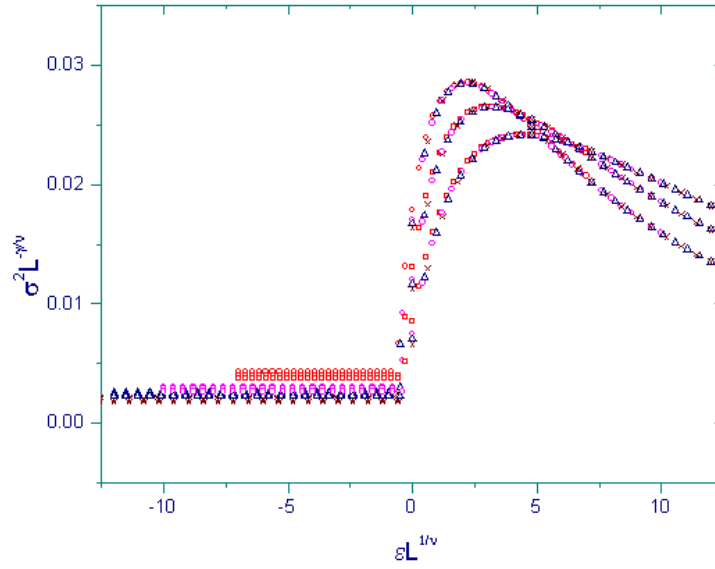


Figura 2.5: Desvio quadrático médio reescalado como função da probabilidade de replicação exata reescalada. Os parâmetros são $1/\nu = 1$, $\gamma/\nu = 1.96$, e (de cima para baixo na região do pico) $a = 10, 20$ e 50 . A convenção é a mesma utilizada na Figura 2.2.

pode ser eliminada por meio da definição de novas variáveis que incorporam a . Este procedimento está ilustrado na Fig. 2.6, onde exibimos a função d versus uma nova variável, definida de tal modo que a função de escala independe de a . Entretanto, o mesmo êxito não foi possível para as curvas da Fig. 2.5. A razão para isto deve-se à necessidade de incorporação de todas as potências de a para os termos da reescala.

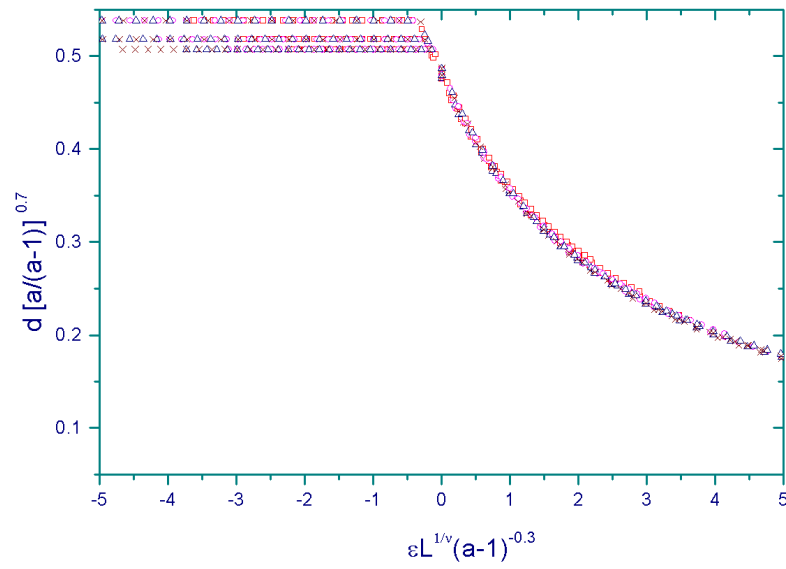


Figura 2.6: Colapso das funções dependentes de a ilustradas na Fig. 2.4. A convenção e os parâmetros são os mesmos para a Fig. 2.2.

Capítulo 3

Uma abordagem para população finita

Até agora tratamos seleção como um processo determinístico. As equações fenomenológicas especificam claramente que cópia entre uma dada população será selecionada. Quando um mutante com vantagem seletiva ($W_{ii} > \bar{E}$) surge, este mutante inevitavelmente dominará a distribuição de mutantes. Existem duas limitações importantes para essa descrição determinística:

- (1) O processo elementar que leva a um mutante específico é inerentemente não-determinístico. A amplificação autocatalítica leva a um mapeamento macroscópico de eventos microscópicos *incertos*.
- (2) O próprio processo de crescimento está sujeito a flutuações estatísticas. Desde que o crescimento inicia-se a partir de uma sequência única, tais flutuações devem ser levadas em consideração. Elas devem modificar apreciavelmente os resultados da teoria determinística, que apenas torna-se válida para um grande número de elementos envolvidos.

Há uma dificuldade adicional surgindo do fato que certos estados estacionários, em contraste ao verdadeiro equilíbrio, são metaestáveis. Devido a todos esses fatos é que devemos reexaminar o problema de seleção do ponto de vista da teoria de probabilidade. Será vis-

to que modificações importantes da teoria fenomenológica resultarão de um tratamento estocástico.

3.1 Descrição do modelo

Nesta seção apresentamos um modelo para evolução de população finita com a função seletiva de pico único (Eq. 2.8) [39]. Para essa função em especial, podemos realizar uma abordagem bastante simples dos efeitos de tamanho finito de população no sistema. A simplicidade vem do aspecto de que a estratificação da população pode ser feita de modo ainda mais compacto que aquela utilizada na seção anterior, baseada na definição de classes de acordo com a distância de Hamming em relação à sequência $(0, 0, \dots, 0)$. Nesta análise, a população de seqüências será dividida em apenas duas classes: uma constituída por seqüências mestras apenas, que aqui denominamos de classe B , e outra constituída por todos os tipos de seqüências diferentes de $\vec{S}_0$, que denominamos de classe C . Portanto a distribuição de valores seletivos será: $A = a$ para os elementos da classe B , e $A = 1$ para os elementos da classe C . Similarmente à descrição dada para o modelo de quase-espécies, cada indivíduo da população é representado por uma seqüência de tamanho L . Porém, aqui nos restringimos apenas ao limite de seqüência infinita, $L \rightarrow \infty$. Este limite apresenta um aspecto interessante, pois podemos desprezar as mutações provenientes dos elementos da classe C para a classe B (mutações reversas), que contribuiriam assim para um aumento da população de seqüências mestras. Podemos justificar facilmente esta simplificação utilizando a definição dada na seção precedente para a matriz de replicação M (Eq. 2.7). Como um exemplo consideremos o elemento M_{01} que, nos denota a probabilidade de criação de uma seqüência mestra por meio da replicação errônea de um mutante de classe 1 e, é dado por

$$M_{01} = q^{L-1}(1 - q). \quad (3.1)$$

Utilizando a igualdade

$$q^L = Q \implies q = Q^{1/L} = e^{\frac{\ln Q}{L}}, \quad (3.2)$$

podemos reescrever a Eq. (3.1) como

$$M_{01} = \frac{Q}{e^{\frac{\ln Q}{L}}} (1 - e^{\frac{\ln Q}{L}}). \quad (3.3)$$

No limite $L \rightarrow \infty$ podemos escrever

$$e^{\frac{\ln Q}{L}} \approx 1 + \frac{1}{L} \ln Q, \quad (3.4)$$

de modo que

$$M_{01} = Q \left(-\frac{1}{L} \ln Q \right) = -\frac{Q}{L} \ln Q \xrightarrow{L \rightarrow \infty} 0. \quad (3.5)$$

Para mutantes de classes superiores a dependência com $(1/L)$ é ainda de ordem mais elevada, logo não existem transições de seqüências pertencentes à outras classes colaborando para o aumento da população de seqüências mestras. Já mutações no sentido inverso são permitidas.

3.2 Dinâmica do modelo

Em nosso estudo consideramos uma população fixa de tamanho N cuja estrutura é dividida em duas classes apenas, uma composta apenas por seqüências mestras, e a outra composta pelos demais tipos de seqüências, porém de sub-estrutura indeterminada. A cada geração a população presente dá origem a uma nova população de mesmo tamanho, sendo posteriormente eliminada (não existe sobreposição de gerações). A formação dessa nova geração é estocástica, sendo o processo de reprodução facilmente implementado por meio da regra da roleta [40]. Conforme pode ser visualizado na Figura 3.1, temos uma roleta que está dividida em duas áreas: a área azul, proporcional ao número de seqüências mestras na população ponderada pelo seu valor seletivo a ; e a área vermelha, proporcional ao número de elementos da classe C e também é ponderada pelo seu valor seletivo, no caso

igual a 1. No processo de reprodução também admitimos a possibilidade de ocorrência de mutações, embora essas mutações ocorram sempre da classe de seqüências mestras para a classe composta pelos demais mutantes.

A dinâmica do nosso sistema é então definida da seguinte forma: a classe de cada indivíduo na próxima geração é escolhida aleatoriamente por meio da regra da roleta, ou seja, a probabilidade de que esse novo indivíduo pertença a uma determinada classe é proporcional à área que essa classe ocupa na roleta. A probabilidade de que o novo indivíduo pertença à classe B é denotada por ω_m e, conseqüentemente, a probabilidade de que esse indivíduo pertença à classe C é igual a $1 - \omega_m$. Se essa seqüência pertencer à classe de seqüências mestras ela tem uma chance de mutar para a C igual a $(1 - Q)$, onde Q determina o fator de qualidade do processo de replicação.

Com a definição dos valores seletivos independentes das freqüências das classes na população e sem a possibilidade de recombinação entre as seqüências, mutação e seleção são forças opostas. Em geral, a seleção tende a aumentar a freqüência do genótipo com o maior valor seletivo, enquanto a mutação tende a reduzi-lo. É intuitivamente óbvio que deva existir um limite superior para a taxa de mutação e que possa ser tolerado pela população. Além deste limite, a seleção não consegue mais compensar o afluxo de mutantes de baixo valor seletivo no sistema, e o processo de mutação torna-se a força dominante governando a composição da população, resultando na inevitável deterioração genética. Müller [41] observou que no caso de uma população assexuada ocorre uma redução aleatória progressiva dos indivíduos mais adaptados, um processo denominado por Felsenstein [3] de *catraca de Müller*. Como resultado temos então uma redução do valor seletivo médio da população devido à acumulação de mutações deletéreas, processo este denominado de *carga mutacional* [42].

Em nossa abordagem, em que a população está dividida em apenas duas classes distintas, o sistema apresentará sempre como estado assintótico o estado caracterizado

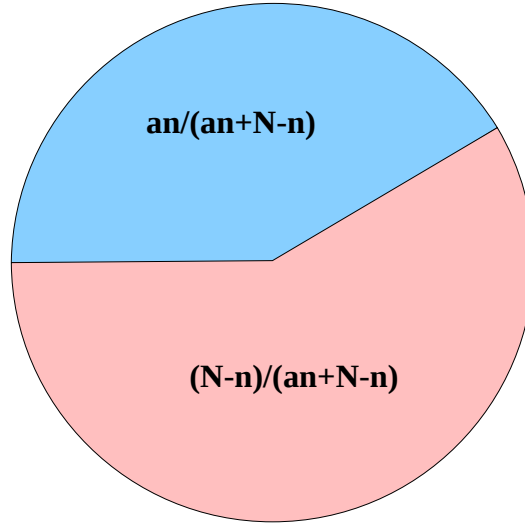


Figura 3.1: A área azul representa a fração de seqüências mestras na população ponderada pelo seu valor adaptativo a , e a área vermelha a fração da população composta pelas demais classes.

por $n = 0$, onde n denota o número de seqüências mestras na população [43]. Isso é esperado pois as flutuações irão certamente eliminar todas as seqüências mestras e não existe dispositivo capaz de restaurá-las na população, pois mutações reversas não ocorrem no limite $L \rightarrow \infty$. De particular interesse nesse problema é a determinação do tempo de relaxação τ do sistema. O tempo de relaxação é definido como o tempo necessário para que o número de seqüências mestras se reduza a um valor $1/e$ de seu número inicial. Como sabemos, este tempo diverge abaixo do limiar de erro $Q = Q_c = 1/a$ para o limite de população infinita $N \rightarrow \infty$. Conforme pode ser observado na Fig. 2.1, o estado estacionário nessa região é caracterizado por uma freqüência não-nula de seqüências mestras. Um aspecto interessante nessa análise é a determinação do comportamento de τ com o número de moléculas N na população. Os resultados numéricos obtidos são baseados na determinação da evolução temporal da distribuição de probabilidades para o número de seqüências mestras na população. O processo dinâmico é descrito por uma cadeia de Markov, desde que nosso problema envolve transições entre valores de uma variável estocástica discreta, ocorrendo em tempos discretos. Cada estado possível do

sistema é caracterizado pelo número de seqüências mestras $n = 0, \dots, N$. Assim, devemos primeiramente introduzir a matriz de transição T , a ser definida a seguir.

3.3 A matriz de transição T

Dada a dinâmica, estamos aptos a obter a matriz de transição $T = \langle T(n, m) \rangle$ do sistema. Os elementos $T(n, m)$ da matriz nos dão a probabilidade de que o sistema transicione de um estado caracterizado por m seqüências mestras na geração atual para um estado caracterizado por n seqüências mestras na geração posterior. De acordo com a dinâmica discutida anteriormente, os elementos de $T(n, m)$ são dados por

$$T(n, m) = \sum_{k=n}^N \binom{N}{k} \binom{k}{n} \omega_m^k (1 - \omega_m)^{N-k} Q^n (1 - Q)^{k-n}, \quad (3.6)$$

onde

$$\omega_m = \frac{am}{am + (N - m)} = \frac{am}{m(a - 1) + N}. \quad (3.7)$$

A primeira etapa da dinâmica, referente ao processo de reprodução, é descrita pelo termo $\binom{N}{k} \omega_m^k (1 - \omega_m)^{N-k}$, que declara a probabilidade de escolhermos k seqüências mestras para a composição da nova população de N indivíduos. A segunda etapa da dinâmica consiste em submeter os indivíduos da classe B ao processo de mutação. Esse processo é descrito pelo termo $\binom{k}{n} Q^n (1 - Q)^{k-n}$, que define a probabilidade de $(k - n)$ indivíduos da classe B mutarem. Desde que estes dois eventos são independentes, os elementos de T são dados pelo produto desses dois termos. A soma sobre k leva em conta todas as possibilidades para se chegar ao estado final, a saber, número de seqüências de mestras igual a n , a partir do dado inicial, número de seqüências mestras igual a m . Como a matriz T define

probabilidades de transição, temos que

$$\sum_{n=0}^N T(n, m) = 1, \quad (3.8)$$

isto é, o sistema deve evoluir para algum estado com o número de seqüências mestras compreendido entre 0 e N .

Podemos agora determinar a evolução temporal do vetor distribuição de probabilidades $\mathbf{P}(t+1)$, onde cada elemento $P_n(t+1)$ representa a probabilidade de encontrarmos $n = 0, \dots, N$ seqüências mestras no instante $t+1$ na população. Essa evolução é dada por

$$\begin{pmatrix} P_0(t+1) \\ P_1(t+1) \\ \vdots \\ P_N(t+1) \end{pmatrix} = \begin{pmatrix} T(0,0) & T(0,1) & \dots & T(0,N) \\ T(1,0) & T(1,1) & \dots & T(1,N) \\ \vdots & \vdots & & \vdots \\ T(N,0) & T(N,1) & \dots & T(N,N) \end{pmatrix} \begin{pmatrix} P_0(t) \\ P_1(t) \\ \vdots \\ P_N(t) \end{pmatrix}. \quad (3.9)$$

Como essa relação é a mesma para todo instante t , e sabendo que $T = \langle T(n, m) \rangle$ é independente do tempo, obtemos portanto

$$\begin{pmatrix} P_0(t) \\ P_1(t) \\ \vdots \\ P_N(t) \end{pmatrix} = \begin{pmatrix} T(0,0) & T(0,1) & \dots & T(0,N) \\ T(1,0) & T(1,1) & \dots & T(1,N) \\ \vdots & \vdots & & \vdots \\ T(N,0) & T(N,1) & \dots & T(N,N) \end{pmatrix}^t \begin{pmatrix} P_0(0) \\ P_1(0) \\ \vdots \\ P_N(0) \end{pmatrix}. \quad (3.10)$$

Ao invés de realizarmos sucessivas operações multiplicativas sobre a matriz T , vamos reduzir enormemente esse cálculo através do seguinte procedimento: Escrevendo o vetor $\mathbf{P}$ no instante $t = 0$ como uma combinação linear de todos os autovetores da matriz T , podemos expressar a evolução temporal apenas em função dos autovalores dessa matriz [44]. Como um exemplo, consideremos o caso $N = 2$. Sejam

$$\begin{pmatrix} l_0 \\ m_0 \\ n_0 \end{pmatrix}, \begin{pmatrix} l_1 \\ m_1 \\ n_1 \end{pmatrix} \quad e \quad \begin{pmatrix} l_2 \\ m_2 \\ n_2 \end{pmatrix} \quad (3.11)$$

autovetores de T , de forma que

$$T \begin{pmatrix} l_i \\ m_i \\ n_i \end{pmatrix} = \lambda_i \begin{pmatrix} l_i \\ m_i \\ n_i \end{pmatrix}, \quad (3.12)$$

onde λ_i denota o i -ésimo autovalor da matriz T . Expressando o vetor $\mathbf{P}(0)$ como combinação linear dos autovetores de T :

$$\mathbf{P}(0) = \begin{pmatrix} P_0(0) \\ P_1(0) \\ P_2(0) \end{pmatrix} = \begin{pmatrix} c_0 l_0 + c_1 l_1 + c_2 l_2 \\ c_0 m_0 + c_1 m_1 + c_2 m_2 \\ c_0 n_0 + c_1 n_1 + c_2 n_2 \end{pmatrix}, \quad (3.13)$$

(se as três equações forem distintas, elas poderão ser resolvidas para c_0, c_1 e c_2) e substituindo a Eq. (3.13) na Eq. (3.10) obtemos

$$\begin{pmatrix} P_0(t) \\ P_1(t) \\ P_2(t) \end{pmatrix} = c_0 T^t \begin{pmatrix} l_0 \\ m_0 \\ n_0 \end{pmatrix} + c_1 T^t \begin{pmatrix} l_1 \\ m_1 \\ n_1 \end{pmatrix} + c_2 T^t \begin{pmatrix} l_2 \\ m_2 \\ n_2 \end{pmatrix}. \quad (3.14)$$

Finalmente utilizando a expressão (3.12), a equação acima resulta em

$$\begin{pmatrix} P_0(t) \\ P_1(t) \\ P_2(t) \end{pmatrix} = c_0 \lambda_0^t \begin{pmatrix} l_0 \\ m_0 \\ n_0 \end{pmatrix} + c_1 \lambda_1^t \begin{pmatrix} l_1 \\ m_1 \\ n_1 \end{pmatrix} + c_2 \lambda_2^t \begin{pmatrix} l_2 \\ m_2 \\ n_2 \end{pmatrix}. \quad (3.15)$$

Logo a evolução temporal da distribuição de probabilidades está totalmente caracterizada pelos autovalores de T .

3.4 Tempo de relaxação

O tempo de relaxação τ é definido como o número de gerações necessárias para que o número médio de seqüências mestras seja reduzido a $1/e$ do seu número inicial.

Da definição dos elementos de matriz $T(n, m)$ dados pela Eq. (3.6), podemos facilmente verificar que a forma apresentada por T é

$$T = \begin{pmatrix} 1 & T(0,1) & . & . & . & T(0,N) \\ 0 & T(1,1) & . & . & . & T(1,N) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & T(N,1) & . & . & . & T(N,N) \end{pmatrix}. \quad (3.16)$$

Como não existem mutações reversas, o único elemento não-nulo da primeira coluna é o elemento $T(0,0)$.

Pelo Teorema de Perron-Frobenius, eis algumas propriedades da matriz de transição T dada a sua forma [44, 45]:

- (i) Possui um autovalor igual a 1 e não-degenerado.
- (ii) Todos os demais autovalores têm módulo menor que 1, $|\lambda| < 1$.
- (iii) Ao autovalor $\lambda = 1$, corresponde um autovetor com componentes não-negativas.

Sejam $\lambda_0, \lambda_1, \dots, \lambda_N$ os autovalores de T tal que $\lambda_0 > \lambda_1 > \dots > \lambda_N$. Para valores grandes de t , temos que λ_0^t é muito maior que $\lambda_1^t, \dots, \lambda_N^t$. Desta forma, depois de muitas gerações a evolução da população passa a ser governada pelo autovalor dominante $\lambda_0 = 1$, cujo autovetor associado é igual a $(1, 0, \dots, 0)$. Ou seja, o estado assintótico que é o estado determinado pelo maior autovalor da matriz ($\lambda_0 = 1$), é dado por $\mathbf{P}(\infty) = (1, 0, \dots, 0)$, conforme mencionado anteriormente.

Generalizando a Eq. (3.15), verificamos que a componente $P_0(t)$ de $\mathbf{P}(t)$ possui a seguinte forma:

$$P_0(t) = \sum_{i=0}^N c_i \lambda_i^t l_i = 1 + c_1 \lambda_1^t l_1 + \dots + c_N \lambda_N^t l_N, \quad (3.17)$$

e apresenta valor de regime estacionário igual a $P_0(\infty) = 1$, desde que $\lambda_i < 1, i = 1, \dots, N$. Como $P_0(t)$ é uma grandeza estritamente positiva e menor ou igual a 1, a soma dos termos com $i \neq 0$ contribui negativamente para Eq. (3.17). Mas como a evolução de $\sum_{i \neq 0} c_i \lambda_i^t l_i$ é determinada pelo autovalor λ_1 , podemos definir um tempo de relaxação τ simplesmente

como o tempo necessário para que $\sum_{i \neq 0} c_i \lambda_i^t l_i$ ou, de acordo com nossa suposição, λ_1^t se reduza a um valor $1/e$ do seu valor inicial. Logo, a evolução temporal será governada pelo termo

$$\lambda_1^t = e^{\ln \lambda_1^t} = e^{t \ln \lambda_1} = e^{-t |\ln \lambda_1|}, \quad (3.18)$$

pois $\ln \lambda_1 < 0$. Assim, podemos definir $\tau \equiv 1/|\ln \lambda_1|$.

Para a determinação do segundo maior autovalor λ_1 , utilizamos o método das potências [46] (Ver apêndice C). Esse método consiste de um procedimento iterativo bastante simples, que tem como objetivo a determinação dos autovalores dominantes de uma matriz sem a necessidade da determinação dos demais autovalores e de seus respectivos autovetores. Tal procedimento garante eficiência e rapidez na obtenção dos resultados numéricos comparados com os demais métodos existentes na literatura. Sem essa eficiência seria bastante difícil a obtenção de resultados para tamanhos de população relativamente grandes ($N \approx 600$).

3.5 Cálculo dos momentos

O primeiro momento da distribuição de seqüências mestras evolui no tempo de acordo com

$$\langle n \rangle_{t+1} = \sum_n \sum_m n T(n, m) P_m(t), \quad (3.19)$$

onde os elementos $T(n, m)$ são definidos pela Eq. (3.6). Substituindo esses elementos na equação acima obtemos

$$\langle n \rangle_{t+1} = \sum_n \sum_m \sum_{k=n}^N n P_m(t) \binom{N}{k} \binom{k}{n} \omega_m^k (1 - \omega_m)^{N-k} Q^n (1 - Q)^{k-n}. \quad (3.20)$$

As somas sobre n e k podem ser facilmente efetuadas resultando em

$$\langle n \rangle_{t+1} = QN \langle \omega_m \rangle_t. \quad (3.21)$$

O segundo momento da distribuição de seqüências mestras é definido por

$$\langle n^2 \rangle_{t+1} = \sum_n \sum_m n^2 T(n, m) P_m(t). \quad (3.22)$$

Substituindo mais uma vez $T(n, m)$, obtemos

$$\langle n^2 \rangle_{t+1} = \sum_n \sum_m \sum_{k=n}^N n^2 P_m(t) \binom{N}{k} \binom{k}{n} \omega_m^k (1 - \omega_m)^{N-k} Q^n (1 - Q)^{k-n}, \quad (3.23)$$

e efetuando as somas sobre k e n ,

$$\langle n^2 \rangle_{t+1} = NQ \langle \omega_m \rangle_t + N(N-1)Q^2 \langle \omega_m^2 \rangle_t. \quad (3.24)$$

Devemos agora calcular $\langle \omega_m \rangle_t$, que de acordo com a Eq. (3.7) é igual a

$$\begin{aligned} \langle \omega_m \rangle_t &= \sum_m P_m(t) \frac{am}{m(a-1) + N} \\ &= a \int_0^\infty dx e^{-xN} \sum_m e^{-x[m(a-1)]} P_m(t) m. \end{aligned} \quad (3.25)$$

Expandindo o termo exponencial escrevemos

$$\begin{aligned} \langle \omega_m \rangle_t &= a \int_0^\infty dx e^{-xN} \sum_m \sum_{k=0}^\infty \frac{(1-a)^k x^k}{k!} m^{k+1} P_m(t) \\ &= a \sum_{k=0}^\infty \frac{(1-a)^k}{k!} \langle m^{k+1} \rangle_t \int_0^\infty dx e^{-xN} x^k \\ &= a \sum_{k=0}^\infty \frac{(1-a)^k}{k!} \langle m^{k+1} \rangle_t \frac{I_k}{N^{k+1}}, \end{aligned} \quad (3.26)$$

onde

$$I_k = \int_0^\infty dy e^{-y} y^k. \quad (3.27)$$

Utilizando o mesmo procedimento para calcular $\langle \omega_m^2 \rangle_t$ encontramos

$$\langle \omega_m^2 \rangle_t = a^2 \sum_{k=0}^\infty \frac{(1-a)^k}{k!} \langle m^{k+2} \rangle_t \frac{I_{k+1}}{N^{k+2}}. \quad (3.28)$$

Finalmente mantendo termos de ordem menor ou igual a $1/N$, obtemos

$$\langle n \rangle_{t+1} = Qa \langle n \rangle_t - Qa(a-1) \frac{\langle n^2 \rangle_t}{N} \quad (3.29)$$

e

$$\begin{aligned} \langle n^2 \rangle_{t+1} &= Qa \langle n \rangle_t - Qa(a-1) \frac{\langle n^2 \rangle_t}{N} \\ &\quad + a^2 Q^2 \langle n^2 \rangle_t \left(1 - \frac{1}{N}\right) - 2a^2 Q^2 (a-1) \frac{\langle n^3 \rangle_t}{N}. \end{aligned} \quad (3.30)$$

Como podemos verificar, as equações acima apresentadas não são fechadas, ou seja, existe sempre uma dependência com o momento de ordem imediatamente superior.

3.6 Resultados

Nas Figuras 3.2 e 3.3 exibimos a dependência de τ com Q para valores de parâmetro $a = 2$ e $a = 20$, respectivamente. Para cada valor de a foram consideradas populações de tamanho $N = 100, 200, 300, 500$ e 600 . Observamos que existe uma tendência nítida do crescimento da derivada de τ com o tamanho da população. Esse aumento é contínuo até o limite de $N \rightarrow \infty$ onde ocorre a divergência de τ no limiar de erro $Q_c = 1/a$. A curva sólida mostrada nessas figuras é o resultado analítico do cálculo de τ no limite $N \rightarrow \infty$. Mais especificamente, neste limite a Eq. (3.29) é escrita como

$$\langle n \rangle_{t+1} = Qa \langle n \rangle_t, \quad (3.31)$$

e, portanto, $\langle n \rangle_t = (Qa)^t \langle n \rangle_0$. Daí

$$\tau = -\frac{1}{\ln(Qa)}, \quad (3.32)$$

que apresenta uma divergência em $Q = Q_c = 1/a$. Este é um resultado importante pois indica que no limite $N \rightarrow \infty$ existe uma região $Q > Q_c$ onde a frequência de seqüências

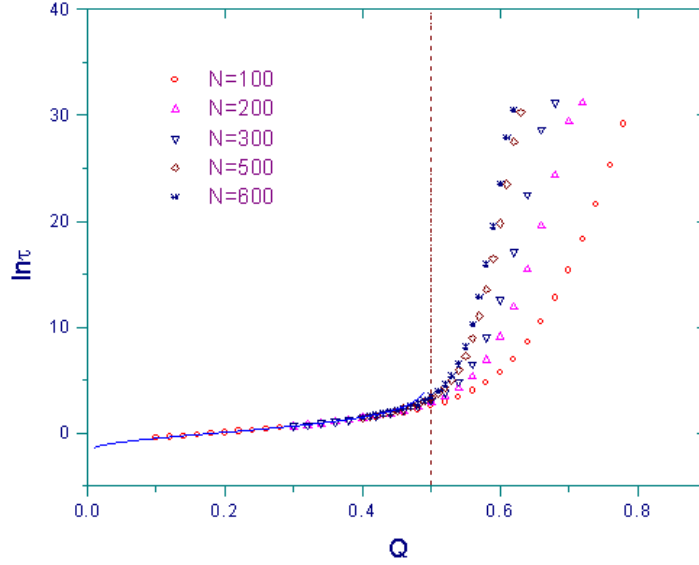


Figura 3.2: Comportamento do tempo de relaxação τ como função de Q para vários valores de N . Neste gráfico o valor seletivo da classe de mestras é $a = 2$. A curva sólida representa o resultado analítico dado pela Eq. (3.32), isto é, $\tau = -\frac{1}{\ln(2Q)}$.

mestras é não-nula no estado estacionário. Isso equivale a dizer que o tempo de relaxação do sistema, como também o tempo de extinção das seqüências mestras, é infinito nesta região.

Nas Figuras 3.4 e 3.5, utilizamos os dados apresentados nas Figuras 3.2 e 3.3, porém agora como função de uma variável reescalada $\zeta = (Q - Q_c)N^{-1/\nu}$, onde $Q_c = 1/a$ e ν é um expoente crítico que determina a dependência da largura da transição com o tamanho N da população, ou seja, o intervalo de valores de Q em torno de Q_c sobre o qual as características da transição persistem. Aplicando mais uma vez o processo adotado anteriormente, podemos definir

$$\tau(Q, N) = N^{-y} f_a(\zeta), \quad (3.33)$$

onde f_a é uma função de escala que é independente do tamanho da população N .

Mais precisamente, verificamos que

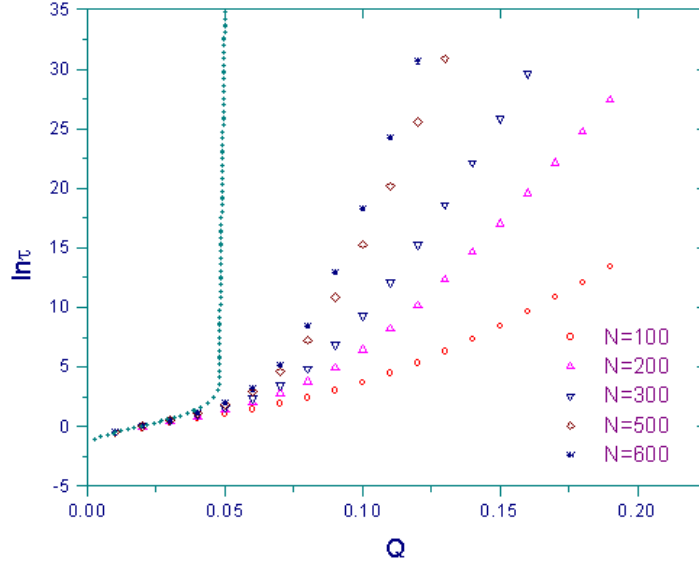


Figura 3.3: Tempo de relaxação τ versus Q para vários valores de N . Neste gráfico o valor seletivo da classe de mestras é $a = 20$. A curva sólida representa o resultado analítico dado pela Eq. (3.32), isto é, $\tau = -\frac{1}{\ln(20Q)}$.

$$\tau(Q_c) \sim N^{1/2}, \quad (3.34)$$

para qualquer valor de a , conforme podemos visualizar na Figura 3.4.

Uma vez identificada a reescala de $\ln \tau$ que leva ao colapso dos dados para diferentes valores de N em $Q = Q_c$, o próximo passo é determinar a largura da transição. Estamos interessados em determinar o valor para o expoente ν , pois introduzimos uma reescala também no eixo ordenado, de forma que na região crítica os dados colapsem. Na Figura 3.5 apresentamos os resultados para a nova função de escala que independe de N . Como podemos observar, obtemos um colapso satisfatório dos dados na região estritamente próxima a $Q = Q_c$ para todos os valores de a . O valor obtido, $1/\nu = 1/2$, independe do parâmetro a e, portanto, exibe uma universalidade. Desta forma, os resultados mostrados indicam que a dependência de τ com N na região crítica é bem descrita pela lei de escala

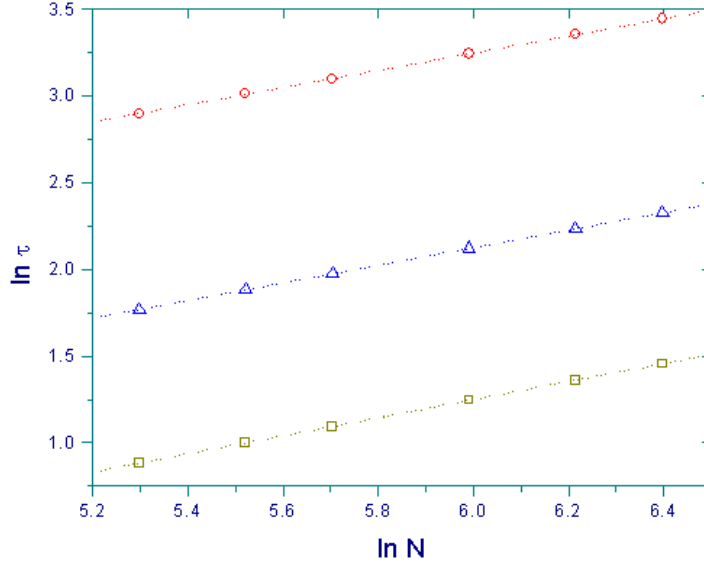


Figura 3.4: Logaritmo do tempo de relaxação, $\ln \tau$, calculado em $Q = Q_c = 1/a$ como função do logaritmo do tamanho da população, $\ln N$, para valores de $a = 2(\bigcirc)$, $a = 10(\triangle)$ e $a = 50(\square)$.

$$\tau = N^{1/2} f_a [(Q - Q_c) N^{1/2}], \quad (3.35)$$

onde f_a é uma função de escala cuja forma específica depende de a .

Este valor para o expoente $1/\nu$ é o mesmo verificado para o modelo de crescimento e morte [30] no limite de $N/a \gg 1$, contradizendo o resultado obtido por Alves et al. [47], onde $\nu = 1$. Da mesma forma, em nosso trabalho verificamos que o colapso dos dados é satisfatório para valores $N \geq 400$ nos casos onde $a = 10$ e 50 . Portanto, de forma a evitarmos erros na estimativa do expoente ν , devemos restringir o procedimento de escalonamento de tamanho finito para os casos onde $N \gg a$.

3.6.1 O limite $a \rightarrow \infty$

O caso limite $a \rightarrow \infty$ apresenta simplificações que nos permitem obter um valor analítico exato para o tempo de relaxação τ .

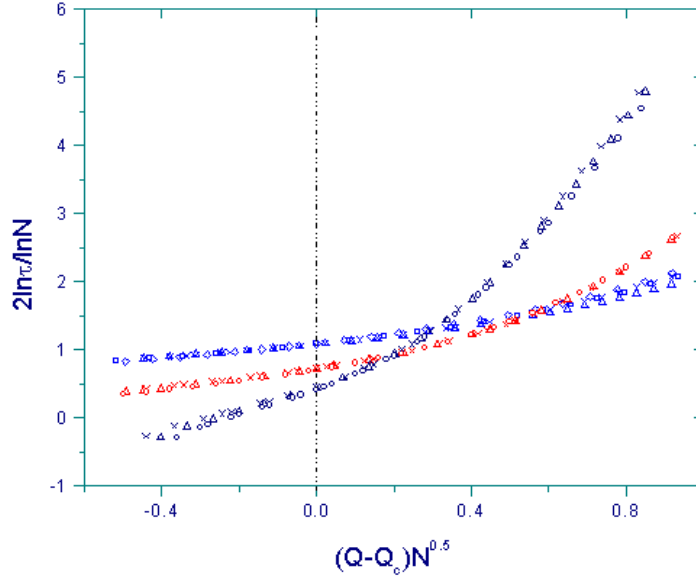


Figura 3.5: Reescala do logaritmo do tempo de relaxação $\ln \tau / \ln N^{1/2}$ como função de $(Q - Q_c)N^{1/2}$ para valores de (de cima para baixo em $Q = Q_c$) $a = 2, 10$ e 50 . A convenção adotada é $N = 200(\diamond), 300(\square), 400(\bigcirc), 500(\triangle)$, e $600(\times)$. Para $a = 10$ e 50 apenas os dados para $N \geq 400$ estão presentes.

A dependência dos elementos de matriz $T(n, m)$ [Eq.(3.6)] com a está contida no termo ω_m , que de acordo com a Eq. (3.7) é igual a

$$\omega_m = \frac{am}{am + (N - m)} \xrightarrow{a \rightarrow \infty} 1. \quad (3.36)$$

Assim, neste limite o único termo que contribui para $T(n, m)$ é $k = N$ e, portanto

$$T(n, m) = T(n) = \binom{N}{n} Q^n (1 - Q)^{N-n}. \quad (3.37)$$

Como podemos verificar, os elementos de T independem de m , ou seja, do estado onde o sistema está no tempo presente. A forma apresentada por T neste caso limite é então,

$$T = \begin{pmatrix} 1 & T(0) & \cdots & T(0) \\ 0 & T(1) & \cdots & T(1) \\ \vdots & \vdots & \vdots & \vdots \\ 0 & T(n) & \cdots & T(n) \end{pmatrix}. \quad (3.38)$$

A solução para o problema de autovalores de uma matriz da forma acima é bastante simples. Os autovalores λ_i de T são iguais a $\lambda=1, [T(1)+\cdots+T(n)], 0, \dots, 0$. Daí,

$$\lambda_1 = \sum_{n=1}^{\infty} \binom{N}{n} Q^n (1-Q)^{N-n} = 1 - (1-Q)^N \quad (3.39)$$

e portanto,

$$\tau = -\frac{1}{\ln [1 - (1-Q)^N]}. \quad (3.40)$$

3.7 Distribuição semi-gaussiana para o número de seqüências mestras

Na seção anterior, encontramos os seguintes resultados para o primeiro e segundo momentos do número de seqüências mestras:

$$\langle n \rangle_{t+1} = NQ \langle \omega_m \rangle_t \quad (3.41)$$

e

$$\langle n^2 \rangle_{t+1} = NQ \langle \omega_m \rangle_t + N(N-1)Q^2 \langle \omega_m^2 \rangle_t, \quad (3.42)$$

onde

$$\langle \omega_m^k \rangle_t = \sum_m P_t(m) \left[\frac{am}{m(a-1) + N} \right]^k. \quad (3.43)$$

No limite $N \rightarrow \infty$, podemos afirmar que no estado estacionário

$$P_t(m) = \delta_{m, n_\infty}, \quad (3.44)$$

onde n_∞ é obtido substituindo a Eq. (3.44) na Eq. (3.41),

$$y_\infty = \frac{Qa-1}{a-1}, \quad (3.45)$$

com $y_\infty = n_\infty/N$. Naturalmente, esta solução é a mesma verificada para a concentração de seqüências mestras no modelo de quase-espécies para o limite de $L \rightarrow \infty$.

Para o limite de seqüência infinita ($L \rightarrow \infty$) e população finita, sabemos que o estado estacionário é caracterizado pelo número de seqüências mestras igual a 0. Portanto, nesse estado a função distribuição no regime estacionário é da forma $P_m(\infty) = \delta_{m,0}$, apresentando variância nula $\sigma^2 = 0$. Entretanto, verificamos que existe um atrator metaestável no qual o número de seqüências mestras é não-nulo. Somente depois de um longo período de tempo ($t \rightarrow \infty$) essas seqüências mestras são eliminadas por flutuações, levando ao resultado esperado.

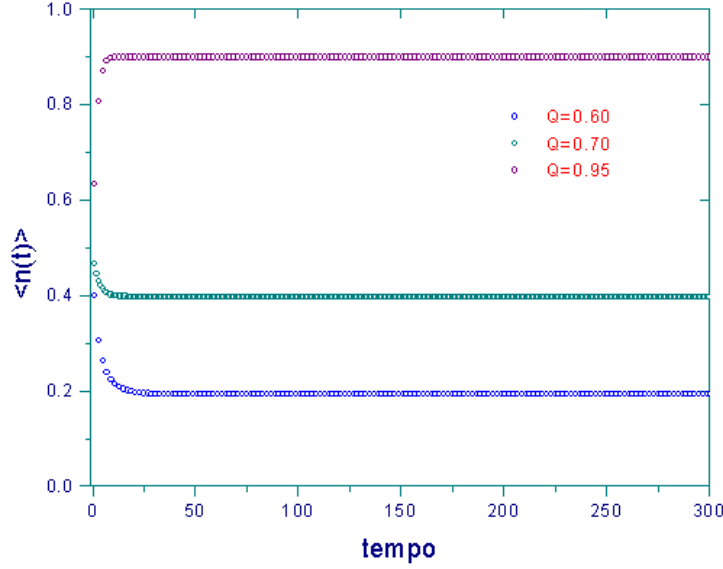


Figura 3.6: Evolução exata da concentração de seqüências mestras no tempo. Os parâmetros utilizados são $a = 2$ e $N = 300$.

Conforme pode ser observado na Figura 3.6, o sistema sai do estado inicial e se desloca rapidamente para o atrator metaestável, que é aproximadamente igual ao valor da concentração de seqüências mestras no limite de $N \rightarrow \infty$ (Eq. 3.45), diferindo de uma pequena quantidade δ que é dependente de N . Seja $\bar{y}$ a fração média de seqüências mestras no atrator para uma população finita N , e σ^2 a variância da distribuição em torno desse valor. Podemos então supor que no regime do atrator

$$\bar{y} = y_\infty + \delta \quad (3.46)$$

e

$$\sigma^2 \neq 0. \quad (3.47)$$

As suposições acima são plausíveis, pois no regime não-estacionário temos $P_m(t) \neq \delta_{m,0}$ e, portanto, existe uma variância em torno do valor médio. Nas Figuras 3.7 e 3.8 exibimos os comportamentos de δ e σ^2 contra Q para valor de parâmetro $a = 10$ e alguns valores de N . Na região estritamente próxima a $Q = Q_c$ o sistema se desloca rapidamente para

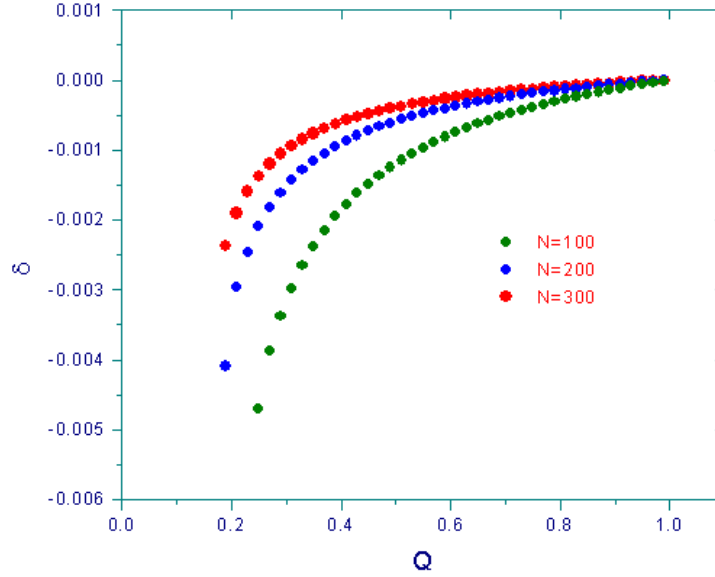


Figura 3.7: δ como função de Q para alguns valores de N . Neste gráfico temos $a = 10$.
o estado assintótico $\bar{y} = 0$ (ver Figura 3.9).

Nosso objetivo nesse estudo é encontrar um valor analítico aproximado para δ e σ^2 . Para tanto, supomos uma distribuição semi-gaussiana para a distribuição de seqüências mestras no tempo, ou seja,

$$P_t(m) = P(y) = \begin{cases} 0 & \text{para } y < 0 \\ \frac{1}{C} e^{-\frac{(y-\bar{y})^2}{2\sigma^2}} & \text{para } y > 0 \end{cases}, \quad (3.48)$$

onde C é a constante de normalização para $P(y)$. Trabalhando com a variável normalizada $y = \frac{n}{N}$ e considerando N relativamente grande, de forma que y pode ser tomada como uma variável contínua, podemos substituir a soma em m na Eq. (3.41) por uma integral em y . Neste caso, a Eq. (3.41) pode ser reescrita como

$$\bar{y} = Qa \int_0^\infty dy P(y) \frac{y}{1 + (a-1)y}, \quad (3.49)$$

com $P(y)$ definida pela Eq. (3.48). Substituindo $P(y)$ temos que,

$$\bar{y} = \frac{Qa}{C} \int_0^\infty dy e^{-\frac{(y-\bar{y})^2}{2\sigma^2}} \frac{y}{1 + (a-1)y}. \quad (3.50)$$

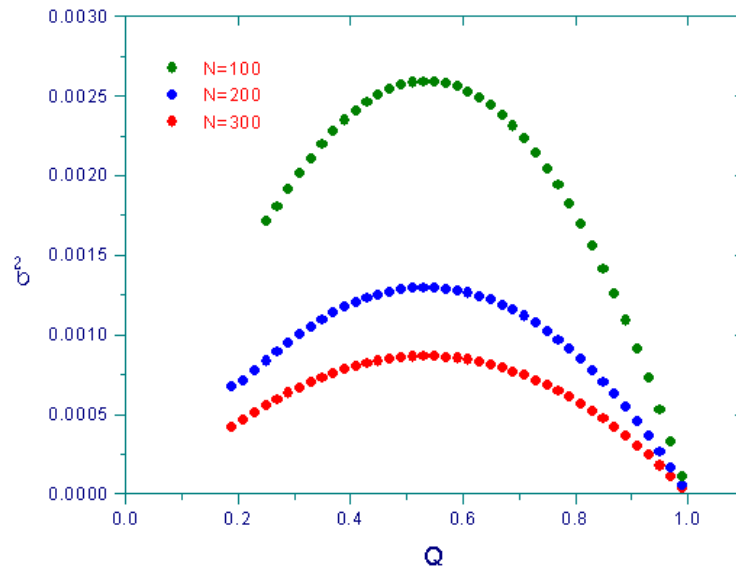


Figura 3.8: σ^2 como função de Q para alguns valores de N . Neste gráfico temos $a = 10$.

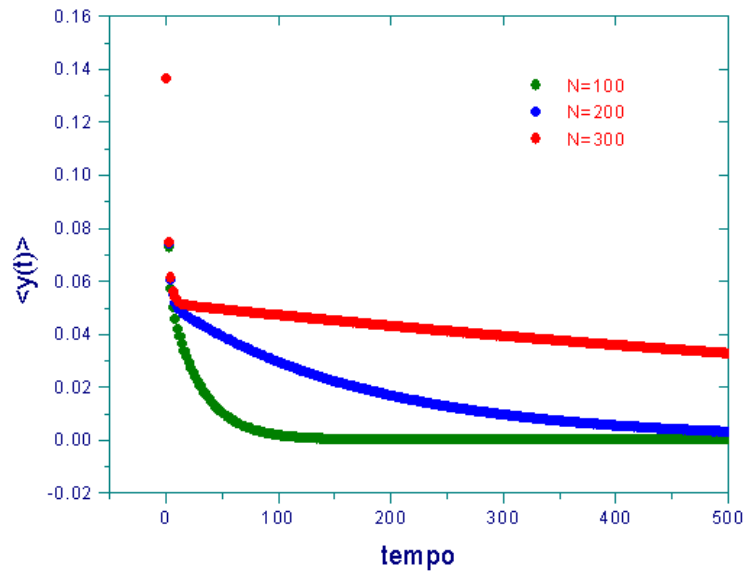


Figura 3.9: Concentração média do número de mestras $\langle y(t) \rangle$ versus *tempo* para alguns valores de N . Como parâmetros temos $a = 10$ e $Q = 0.15$. A transição para $N \rightarrow \infty$ ocorre em $Q_c = 0.1$.

Definindo $z = \frac{y - \bar{y}}{\sqrt{\sigma^2}}$, $\tilde{a} = a - 1$ e tomando o limite de $\varepsilon \equiv \sqrt{\sigma^2} \rightarrow 0$, escrevemos

$$\bar{y} = \frac{Qa}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dz e^{-\frac{z^2}{2}} \frac{(z\varepsilon + \bar{y})}{1 + \tilde{a}(z\varepsilon + \bar{y})}, \quad (3.51)$$

onde usamos $C = \sqrt{2\pi\sigma^2}$. Realizando a substituição $Dz = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$ e reorganizando a equação acima obtemos

$$\bar{y} = Qa \int_{-\infty}^{\infty} Dz \frac{(z\varepsilon + \bar{y})}{1 + \tilde{a}\bar{y}} \frac{1}{\left[1 + \frac{\tilde{a}z\varepsilon}{1 + \tilde{a}\bar{y}}\right]}. \quad (3.52)$$

Efetando agora uma expansão em ε , e mantendo termos até a ordem ε^4 , encontramos

$$\bar{y} = Qa \left\{ \frac{\bar{y}}{1 + \tilde{a}\bar{y}} \left[1 + \frac{\varepsilon^2 \tilde{a}^2}{(1 + \tilde{a}\bar{y})^2} + 3 \frac{\varepsilon^4 \tilde{a}^4}{(1 + \tilde{a}\bar{y})^4} \right] + \frac{1}{1 + \tilde{a}\bar{y}} \left[-\frac{\varepsilon^2 \tilde{a}}{1 + \tilde{a}\bar{y}} - 3 \frac{\varepsilon^4 \tilde{a}^3}{(1 + \tilde{a}\bar{y})^3} \right] \right\}. \quad (3.53)$$

Utilizando a definição de $\bar{y}$, dada na Eq. (3.46), encontramos que

$$\begin{aligned} y_{\infty} + \delta = & Qa \left\{ \frac{y_{\infty} + \delta}{(1 + \tilde{a}y_{\infty}) \left[1 + \frac{\tilde{a}\delta}{1 + \tilde{a}y_{\infty}} \right]} \left[1 + \frac{\varepsilon^2 \tilde{a}^2}{(1 + \tilde{a}y_{\infty})^2 \left[1 + \frac{\tilde{a}\delta}{1 + \tilde{a}y_{\infty}} \right]^2} + 3 \frac{\varepsilon^4 \tilde{a}^4}{(1 + \tilde{a}y_{\infty})^4 \left[1 + \frac{\tilde{a}\delta}{1 + \tilde{a}y_{\infty}} \right]^4} \right] \right. \\ & \left. + \frac{1}{(1 + \tilde{a}y_{\infty}) \left[1 + \frac{\tilde{a}\delta}{1 + \tilde{a}y_{\infty}} \right]} \left[-\frac{\varepsilon^2 \tilde{a}}{(1 + \tilde{a}y_{\infty}) \left[1 + \frac{\tilde{a}\delta}{1 + \tilde{a}y_{\infty}} \right]} - 3 \frac{\varepsilon^4 \tilde{a}^3}{(1 + \tilde{a}y_{\infty})^3 \left[1 + \frac{\tilde{a}\delta}{1 + \tilde{a}y_{\infty}} \right]^3} \right] \right\} \quad (3.54) \end{aligned}$$

Expandindo agora em δ , e mantendo os termos até a ordem ε^4 , δ^2 e $\delta\varepsilon$, obtemos como equação final

$$\delta \left(-\frac{Qa - 1}{Qa} \right) + \sigma^2 \left(-\frac{\tilde{a}}{Q^2 a^2} \right) + \delta^2 \left(-\frac{\tilde{a}}{Q^2 a^2} \right) + \delta \sigma^2 \left(\frac{3\tilde{a}^2}{Q^3 a^3} \right) + \sigma^4 \left(-\frac{3\tilde{a}^3}{Q^4 a^4} \right) = 0, \quad (3.55)$$

onde substituímos y_{∞} pelo seu valor dado na Eq. (3.45) e usamos a igualdade $\varepsilon^2 = \sigma^2$.

Entretanto, verificamos que mantendo apenas os termos de primeira ordem (δ e σ^2) para a equação acima, obtemos resultados em excelente concordância com os valores numéricos exatos para a razão

$$\delta/\sigma^2 = -\frac{a - 1}{Qa(Qa - 1)}, \quad (3.56)$$

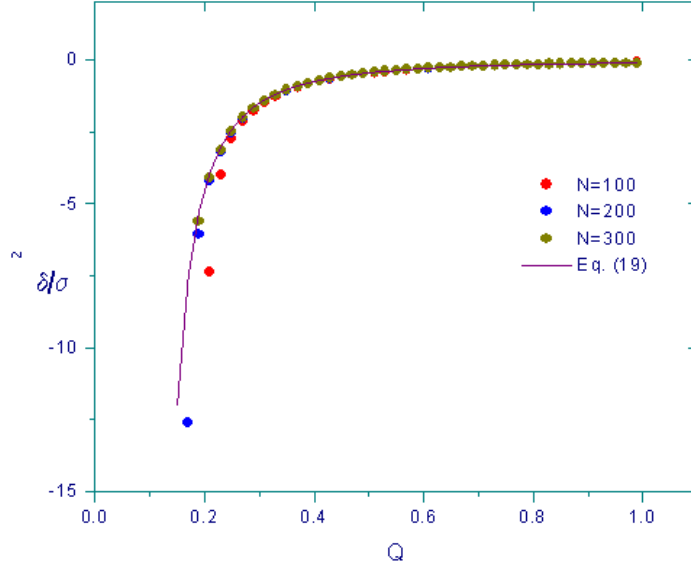


Figura 3.10: Razão δ/σ^2 para valor seletivo $a = 10$ e alguns valores de N . A linha sólida representa o valor teórico aproximado.

conforme ilustrado na Figura 3.10.

Para o segundo momento, dado pela Eq. (3.42), temos que

$$N^2 \overline{y^2} = NQa \int_0^\infty dy P(y) \frac{y}{1 + (a-1)y} + N(N-1)Q^2 a^2 \int_0^\infty dy P(y) \frac{y^2}{[1 + (a-1)y]^2}, \quad (3.57)$$

onde $\overline{y^2} = \overline{y^2} + \sigma^2 = (y_\infty + \delta)^2 + \sigma^2$. Utilizando o mesmo procedimento e aproximações aplicados ao primeiro momento, encontramos como equação final

$$\begin{aligned} & \frac{1}{N} \left(\frac{Qa-1}{a-1} - \frac{(Qa-1)^2}{(a-1)2} \right) + \delta \left(\frac{1}{QaN} - \frac{2(Qa-1)^2}{Qa(a-1)} + \frac{2(Qa-1)^2}{Qa(a-1)N} + \frac{2(Qa-1)}{(a-1)N} \right) \\ & + \delta^2 \left(\frac{-Q^2 a^2 - 2Qa + 3}{Q^2 a^2} \right) + \sigma^2 \left[\left(\frac{-Q^2 a^2 - 2Qa + 3}{Q^2 a^2} \right) \left(1 - \frac{1}{N} \right) + \frac{1}{N} - \frac{a-1}{Q^2 a^2 N} \right] \\ & + \sigma^4 \left(\frac{3\tilde{a}^2(-2Qa+5)}{Q^4 a^4} \right) + \delta \sigma^2 \left(\frac{6\tilde{a}(Qa-2)}{Q^3 a^3} \right) = 0. \end{aligned} \quad (3.58)$$

Utilizando o resultado da Eq. (3.56), podemos determinar a dependência de δ com σ^2 e substituí-la na equação acima. Teremos portanto uma equação quadrática para σ^2 .

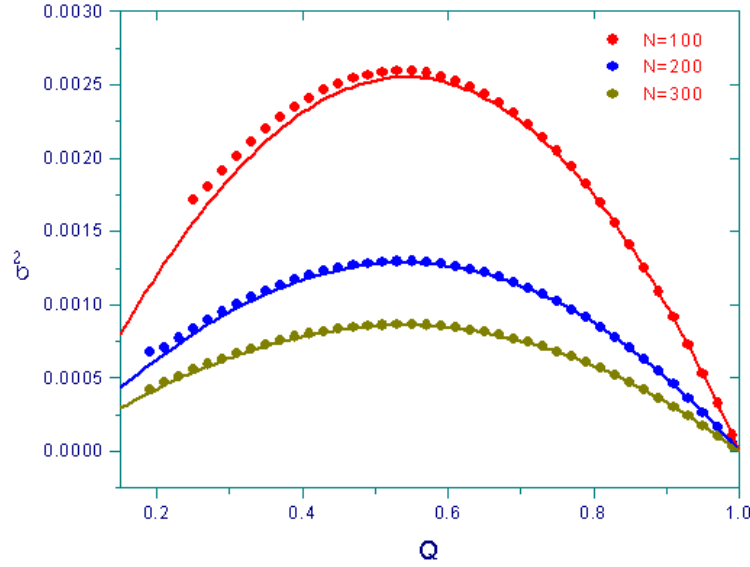


Figura 3.11: σ^2 versus Q para alguns valores de N . Nesta Figura temos $a = 10$. As linhas sólidas representam os valores analíticos de acordo com a distribuição semi-Gaussiana para cada valor de N .

Devemos apenas tomar o cuidado de tomar a raiz que fornece valores sempre positivos para σ^2 na região de interesse.

Observando as Figuras (3.11) e (3.12), constatamos que os valores analíticos para a distribuição semi-gaussiana estão em ótimo acordo com os resultados numéricos exatos. Os resultados numéricos foram obtidos por meio da aplicação sucessiva da matriz de transição T , conforme definições das Eqs. (3.6) e (3.10).

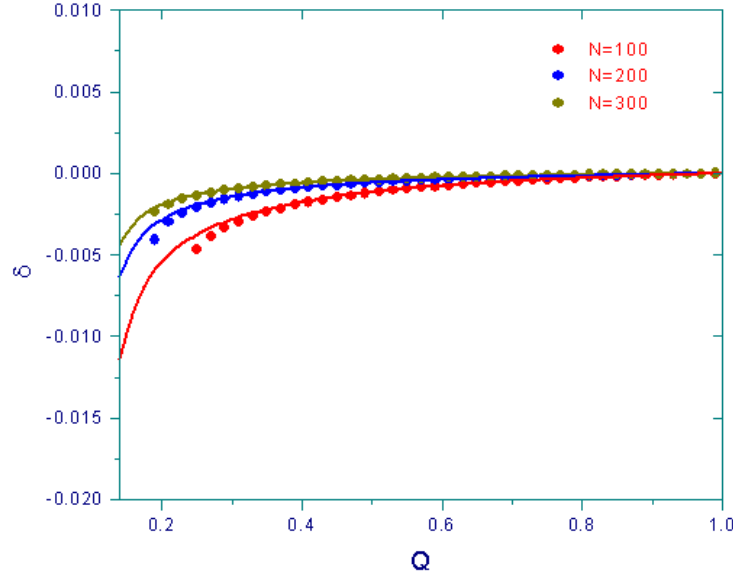


Figura 3.12: δ versus Q para alguns valores de N . Nesta Figura temos $a = 10$. As linhas sólidas representam os valores analíticos de acordo com a distribuição semi-gaussiana para cada valor de N .

3.8 Tempo de extinção

Em nossa abordagem de população finita analisamos a dinâmica do modelo por meio da definição de tempo de relaxação do sistema. Este tempo é definido como o tempo necessário para que a população de seqüências mestras se reduza a um valor igual a $1/e$ do seu valor inicial. Nesta seção, realizamos simulações macroscópicas do modelo em discussão. Nessas simulações a dinâmica do modelo é descrita pela matriz de transição $T = \langle T(n, m) \rangle$, definida na Equação (3.6). Os elementos de T definem as probabilidades de transição do sistema de um estado m em que ele se encontra, para um estado n em um instante imediatamente posterior. Da dinâmica podemos obter o tempo médio de extinção da população de seqüências mestras τ_{ext} , que não é idêntico ao tempo de relaxação τ , como também a distribuição de probabilidade para os eventos de extinção.

O algoritmo é descrito da seguinte forma:

- (1) Iniciamos a população em um estado caracterizado por $m = \frac{N}{2}$ seqüências mestras.
- (2) Definido o estado inicial, realizamos transições conforme os elementos $T(n, m)$, onde $\sum_{n=0}^N T(n, m) = 1$. Para definirmos o estado n' para o qual o sistema irá transicionar, sorteamos um número aleatório r de uma distribuição unitária uniforme e, dividimos esse intervalo em $N + 1$ intervalos distintos de comprimento $T(n, m)$. O estado n' é escolhido de forma que

$$\sum_{n=0}^{n'-1} T(n, m) \leq r \leq \sum_{n=0}^{n'} T(n, m). \quad (3.59)$$

- (3) Repetimos o mesmo procedimento até que a população de mestras seja extinta.

Realizamos diversos experimentos independentes, sempre iniciando de um mesmo estado inicial, de forma a obtermos uma distribuição de probabilidade para o evento de extinção $P_{ext}(t)$ em bom acordo com as previsões numéricas exatas.

A distribuição de probabilidade numérica exata para o evento de extinção é obtida por meio da aplicação iterativa de T ao vetor $\mathbf{P}(t)$, ou seja,

$$\mathbf{P}(t) = T\mathbf{P}(t-1). \quad (3.60)$$

A probabilidade de extinção em um instante t é então dada por

$$P_{ext}(t) = P_0(t) - P_0(t-1), \quad (3.61)$$

onde $P_0(t)$ define a probabilidade de termos $n = 0$ seqüências mestras no instante de tempo t . Daí então, podemos determinar o tempo médio de extinção por meio de

$$\tau_{ext} = \sum_{t=1}^{\infty} t P_{ext}(t). \quad (3.62)$$

Na Figura 3.13 exibimos o resultado numérico exato e também por meio de simulações para a distribuição de probabilidade de evento de extinção em função do tempo. Nesta

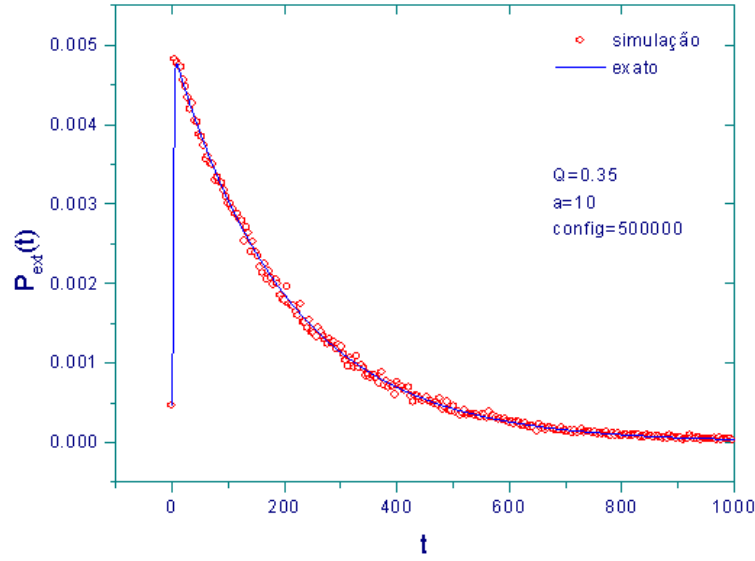


Figura 3.13: Distribuição de probabilidade de evento de extinção das seqüências mestras P_{ext} versus t .

figura temos valores de parâmetro $N = 20$, $a = 10$ e $Q = 0.35$. As médias foram tomadas sobre 500000 configurações. Como podemos observar, existe uma ótima concordância entre os dois resultados.

Na Figura 3.14 apresentamos o comportamento do tempo de extinção τ_{ext} para alguns valores de N . O comportamento é semelhante àquele verificado para τ . Mais uma vez, observamos uma excelente conformidade entre os tempos medidos numericamente através da sucessiva aplicação da matriz T e os tempos obtidos através da simulação.

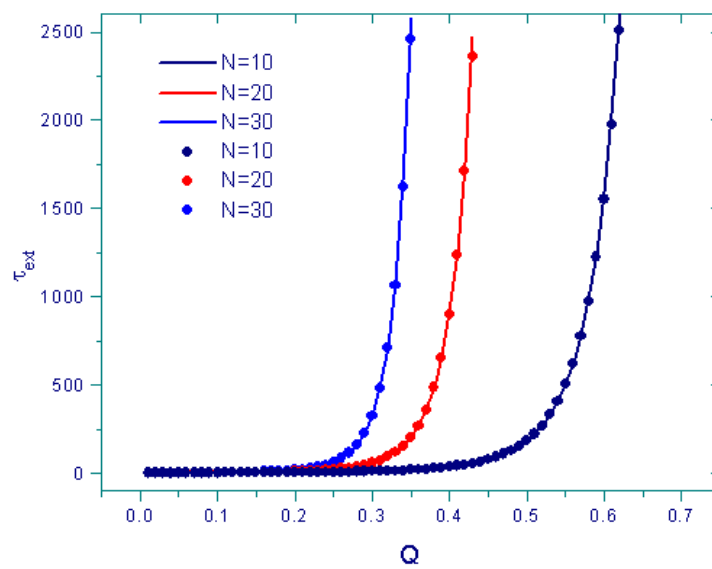


Figura 3.14: Tempo de extinção τ_{ext} versus Q para alguns valores de N . A linha sólida representa os valores numéricos exatos. Os pontos são os dados obtidos por meio de simulação.

Capítulo 4

Hiperciclos

Qualquer que seja um sistema vivo, sua funcionalidade só pode ser concebida em um meio já preparado para aceitar vida. Uma longa evolução química é portanto um pré-requisito necessário para a manifestação de qualquer meio de vida. Entretanto, esse estágio pré-biótico não é o objeto de nosso estudo. O resultado mais impressionante nesse estágio é o que diz respeito à segunda fase da evolução, o surgimento dos primeiros sistemas auto-replicantes. Esse estágio de evolução é certamente o mais complexo, pois compreende a transição de sistemas não-vivos para sistemas com vida.

A questão fundamental é explicar a emergência da ordem e complexidade, que são base do ciclo de auto-reprodução desses sistemas na sopa primitiva de macromoléculas. Muitas tentativas têm sido realizadas para solucionar este problema fundamental em um contexto físico. A teoria física mais completa da origem da vida foi formulada por Eigen [1]. Suas idéias sobre auto-organização da matéria e evolução de macromoléculas biológicas são a base fundamental de nosso trabalho.

As macromoléculas mais importantes sintetizadas por organismos vivos são os ácidos nucléicos, que agem como portadores de informação biológica, e as proteínas, que realizam a maioria das funções biológicas. Esses biopolímeros são os elementos estruturais do ciclo de reprodução. O vínculo entre informação biológica e função biológica é fornecido pelo código genético. *Função* sem instrução (informação) não pode gerar nenhum tipo de

ordem reprodutível, e esta *informação* apenas possui algum significado via a *função* que ela codifica. Por esta razão, é relevante levar em consideração essa interação complexa e altamente coordenada entre ácidos nucléicos e proteínas no estudo de origem da vida. A emergência da evolução e de uma auto-organização funcional convergente a partir desta hipótese de interação entre macromoléculas será alvo de nossa discussão neste capítulo.

Vamos agora nos concentrar no problema mais interessante apresentado pelo modelo de quase-espécies, que é a limitação na quantidade de informação armazenada por uma molécula devido ao limiar de erro (catástrofe de erro). Um outro problema que surge na concepção de quase-espécies moleculares como entidade responsável pela armazenagem da informação, é a constituição desse sistema apenas por mutantes que são biologicamente mais afinizadas com a seqüência mestra (isto é, estruturalmente semelhantes ou, conforme nossa descrição, mais próximos da mestra em termos da distância de Hamming). Essa concepção é um tanto errônea, desde que genes desenvolvendo diferentes funções devem ocupar posições distantes no espaço de seqüências, ou seja, um número relativamente grande de mutações deve existir entre as seqüências que compõem o genoma. Se a seqüência mestra é um dos genes, é pouco provável que os outros genes sejam seqüências estruturalmente similares.

O próprio Eigen esforçou-se na tentativa de resolver essas falhas do modelo de quase-espécies. Em 1978, numa série de três artigos, foi desenvolvido o modelo de hiper-ciclos por Eigen e Schuster [4, 5, 8]. Os hiper-ciclos têm sido postulados como um evento único em evolução pré-biótica e nos dão uma boa explicação para a emergência do código genético, assim como para o aumento do tamanho do genoma. Seu ponto de partida é a observação de que cadeias e ciclos catalíticos são comuns em todos os organismos modernos e, portanto, devem ter evoluído de algum estágio inicial da origem da vida (estágio pré-biótico). Uma cadeia catalítica é uma seqüência de reações químicas onde o produto de uma reação é utilizado como reagente na próxima reação. Um ciclo é meramente uma

cadeia fechada: se existem n reações separadas, o produto da n -ésima reação é necessário como reagente na primeira. Embora Eigen e Schuster incluam uma discussão periférica dos ciclos químicos em geral e uma revisão das idéias atuais nessa problemática como a origem do código genético, a teoria quantitativa concentra-se inteiramente na análise de sistemas de equações diferenciais usadas para descrever tais ciclos catalíticos. O objetivo dessa análise é mostrar que esses ciclos podem permitir a coexistência de várias quase-espécies moleculares. Temos portanto a idéia de que as unidades de nível mais elevado (hiperciclos) devem surgir como o resultado da coevolução de unidades em um nível inferior (quase-espécies).

Eigen e Schuster [4] resumiram os critérios para a estabilidade do hiperciclo da seguinte forma :

- (i) Cada membro da unidade de nível superior deve ser capaz de competir com sucesso com suas cópias errôneas.
- (ii) Apesar de suas tendências competitivas intrínsecas, os membros da unidade devem agir cooperativamente, e não devem por competição excluir uns aos outros.
- (iii) O sistema deve competir com sucesso com os membros menos eficientes ou sistemas inteiros.

Estes são os critérios para a integração da informação genética distribuída em elementos replicados separadamente.

As equações diferenciais estudadas no modelo de hiper ciclos aplicam-se a qualquer sistema cíclico de reações químicas, biológico ou não. Em nosso contexto, devemos visualizar cada espécie química como uma quase-espécie. O modelo de hiper ciclos supõe que uma coleção dessas espécies foram acumuladas e que algum tipo de conexão cíclica foi estabelecida entre duas ou mais delas.

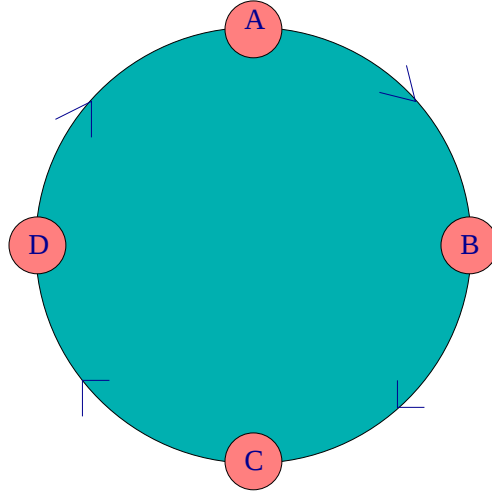


Figura 4.1: O hiperciclo. Cada uma das unidades A, B, C e D é um replicador. A taxa de replicação de cada unidade é uma função crescente da concentração da unidade imediatamente anterior. Assim a taxa de replicação de B é uma função da concentração de A, e assim por diante.

4.1 O modelo

Consideremos um sistema de n espécies, com a frequência da espécie i denotada por $x_i(t)$, $i = 1, \dots, n$. Em geral, a evolução temporal dessas frequências é governada por um sistema de equações diferenciais ordinárias de primeira ordem:

$$\dot{x}_i = \Lambda_i(x_1, \dots, x_n; k_1, \dots, k_n; t), \quad (4.1)$$

onde Λ_i é em geral função não-linear de seus argumentos; os k'_i s são parâmetros e t é o tempo.

Em todos os casos que consideramos, as Eqs. (4.1) são autônomas, isto é, não apresentam uma dependência explícita no tempo. Se definirmos os vetores

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_N \end{pmatrix} \quad e \quad \mathbf{k} = \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ \vdots \\ k_n \end{pmatrix} \quad (4.2)$$

então o sistema de Eqs. (4.1) pode ser escrito como

$$\dot{x}_i = \Lambda_i(\mathbf{x}, \mathbf{k}). \quad (4.3)$$

Devemos dividir as contribuições de cada função Λ_i em duas partes: a primeira emergindo de processos químicos tais como reações ou degradação de moléculas, denotada por Γ_i ; e a segunda consistindo do fluxo de moléculas em ambos os sentidos (de fora para dentro e de dentro para fora do sistema), denotada por ϕ_i . Daí,

$$\dot{x}_i = \Gamma_i + \phi_i. \quad (4.4)$$

Vamos supor que o fluxo da espécie i , ϕ_i , seja proporcional ao fluxo total $\phi(t)$, de forma que

$$\phi_i = \frac{x_i(t)}{c(t)} \phi(t), \quad (4.5)$$

onde $c(t)$ é a concentração de todas as espécies, que anteriormente igualamos a 1, ou seja,

$$c(t) = \sum_{i=1}^n x_i(t). \quad (4.6)$$

A taxa de variação total da população é portanto igual a

$$\dot{c} = \sum_i \dot{x}_i = \sum_i \Gamma_i + \phi. \quad (4.7)$$

Verificamos assim, que devemos incluir algum vínculo no sistema por meio da especificação de uma condição para a concentração total c ou para o fluxo total ϕ , mas não em ambos, pois um é determinado pelo outro. Como um exemplo mais comum na literatura, podemos impor o vínculo de que a concentração total permaneça constante, isto é, $\dot{c} = 0$ e $c(t) = c_0 = cte$. Neste caso,

$$\sum_i \Gamma_i = -\phi \quad (4.8)$$

e

$$\dot{x}_i = \Gamma_i - \frac{x_i}{c_0} \sum_j \Gamma_j. \quad (4.9)$$

No caso geral, a concentração total não precisa ser constante, de forma que a Eq. (4.9) iguala a

$$\dot{x}_i = \Gamma_i + \frac{x_i}{c} \dot{\phi} = \Gamma_i - \frac{x_i}{c} \sum_j \Gamma_j + \frac{x_i}{c} \dot{c}. \quad (4.10)$$

Devemos então introduzir coordenadas normalizadas $\xi_i = x_i/c$ tal que $\sum_i \xi_i = 1$. Em termos dessas coordenadas a Eq. (4.10) é reescrita como

$$\dot{\xi}_i = \frac{1}{c} \left[\Gamma_i(\mathbf{x}) - \xi_i \sum_j \Gamma_j(\mathbf{x}) \right]. \quad (4.11)$$

Vamos considerar agora um exemplo simples de hiperciclo formado por 3 espécies moleculares. Se a taxa de produção de uma molécula da espécie 1 pode ocorrer por meio da interação com uma molécula da espécie 2 (catálise) e também por auto-replicação, então uma forma possível para a função Γ_1 é

$$\Gamma_1 = k_1 x_1 x_2 + k_2 x_1, \quad (4.12)$$

onde k_1 é a taxa com que a reação (espécie 1)-(espécie 2) ocorre, e k_2 a taxa com que a espécie 1 pode se auto-reproduzir. Evidentemente, muitas possibilidades existem para os Γ_i 's, dependendo das particularidades das reações químicas que produzem cada espécie. Uma forma particularmente simples para as equações é obtida se supusermos que todos os Γ_i 's são polinômios homogêneos de grau p dos x_i 's, ou seja, todos os termos em cada Γ_i têm precisamente p fatores de x_i . O exemplo dado acima [Eq.(4.12)] não é homogêneo, desde que contém um termo de grau 1 e outro de grau 2. Um exemplo de um polinômio homogêneo de grau $p = 2$ é

$$\Gamma_i = k_1 x_1 x_2 + k_2 x_2 x_3. \quad (4.13)$$

Essa condição de homogeneidade é equivalente à suposição de que todas as reações que produzem moléculas de qualquer espécie requerem o mesmo número de reagentes. Enquanto isto certamente não é o que ocorre em organismos mais evoluídos, provavelmente

é uma boa aproximação numa situação primitiva onde soluções de tais moléculas seriam muito diluídas. Em tais casos, seria improvável que qualquer reação necessitasse de mais que dois reagentes, uma vez que em condições de diluição extrema a probabilidade de que três ou mais moléculas interajam em uma mesma região num intervalo de tempo relativamente curto é desprezível.

Supondo que todos os Γ'_i s sejam polinômios homogêneos de grau p , a conversão de x_i para ξ_i torna-se

$$\Gamma_i(\mathbf{x}) = c^p \Gamma_i(\xi) \quad (4.14)$$

e as equações diferenciais são escritas como

$$\dot{\xi}_i = c^{p-1} \left[\Gamma_i(\xi) - \xi_i \sum_j \Gamma_j(\xi) \right]. \quad (4.15)$$

A dependência na concentração é agora restrita a um único fator, a saber c^{p-1} . Assim, dado que a concentração permanece finita e não-nula, os pontos fixos do sistema são os mesmos, sendo a concentração constante ou não.

4.2 O espaço de concentrações *Simplex*

As variáveis de concentração x_i , $i = 1, \dots, n$, são grandezas não-negativas: $0 \leq x_i \leq \infty$.

Assim as variáveis normalizadas ξ_i satisfazem as seguintes condições:

$$0 \leq \xi_i \leq 1 \quad e \quad \sum_{i=1}^n \xi_i = 1 \quad i = 1, \dots, n. \quad (4.16)$$

A região ocupada pelo conjunto de valores de $\{\xi_i\}$ é denominada de *simplex*. Geometricamente, um simplex é um poliedro regular de dimensão $n - 1$ com n vértices. Na Fig. 4.2 mostramos um exemplo de simplex com $n = 3$ (triângulo equilátero).

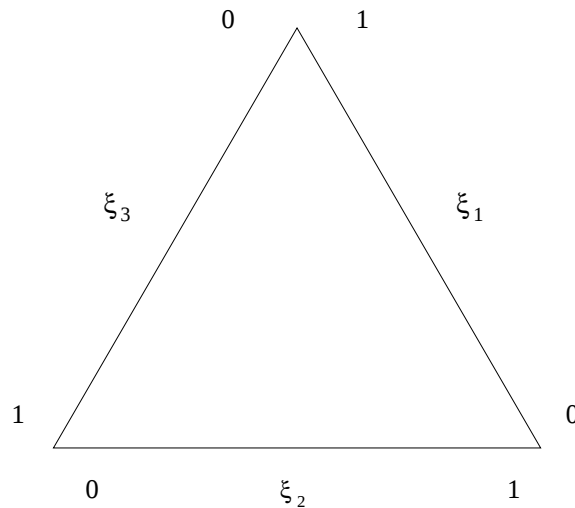


Figura 4.2: Um simplex para 3 espécies.

O simplex está sempre inserido em um espaço de dimensão n associado às verdadeiras variáveis de concentrações x_i . Na Fig. 4.3 exibimos o mesmo simplex com $n = 3$ inserido no espaço tri-dimensional com seus vértices tocando os três eixos de coordenadas. Para determinar o ponto no simplex correspondente a um conjunto particular de valores de concentrações x_i , traçamos uma linha da origem ao ponto representando a configuração de concentração. O ponto onde esta linha intercepta o simplex determina os valores correspondentes das variáveis ξ_i . A representação de um sistema em um simplex, ao invés do próprio espaço n -dimensional, tem a vantagem de nos permitir, por exemplo, acompanhar a evolução de um sistema de três espécies em um diagrama bi-dimensional. Na análise de vários processos moleculares de auto-organização estamos em geral mais interessados no resultado final da evolução do que numa resolução detalhada da dinâmica do processo. Uma análise comparativa do comportamento evolutivo como função dos parâmetros de controle do modelo não requer necessariamente um conhecimento completo das curvas de solução do sistema de equações diferenciais. Assim, a determinação e análise dos pontos fixos é o procedimento comumente escolhido na literatura especializada, pois é o que melhor atende a esses propósitos, principalmente quando o interesse maior é o

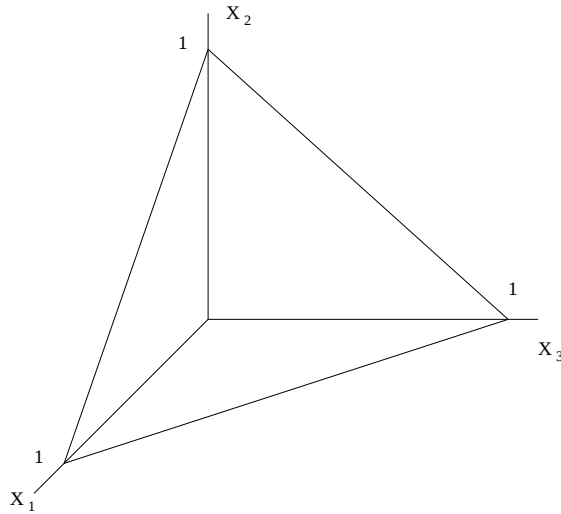


Figura 4.3: O simplex 3-espécies inserido no espaço tri-dimensional.

comportamento de regime estacionário de um sistema dinâmico [48].

Nos casos onde o regime estacionário do sistema independe da concentração total c (isto é, quando as funções Γ_i são homogêneas), os pontos fixos nas coordenadas normalizadas não mudam quando c é alterado. Assim, a real localização de um ponto fixo no espaço x_i pode ser obtida para qualquer tamanho do sistema por meio da determinação do vetor no plano do simplex, que se estende da origem ao ponto fixo no simplex, sendo posteriormente multiplicado por c . Entretanto, sabemos que se os Γ_i não forem funções polinomiais homogêneas nas variáveis de concentração x_i , a localização dos pontos fixos quando o sistema é representado nas coordenadas normalizadas ξ_i não será independente da concentração total c . Logo, não podemos obter os pontos fixos nas variáveis x_i através de uma simples operação de multiplicação e então é mais fácil trabalhar diretamente com as variáveis x_i .

Na Fig. 4.4 exibimos o ponto $[0.25, 0.25, 0.5]$ no simplex, onde os catetos do triângulo correspondem às variáveis normalizadas. A representação de um ponto no simplex é bastante simples. Linhas de ξ_i constante são paralelas ao eixo ξ_{i-1} , onde os índices são cíclicos, isto é, linhas de ξ_1 constante são paralelas ao eixo ξ_3 , linhas de $\xi_2 = cte$ são

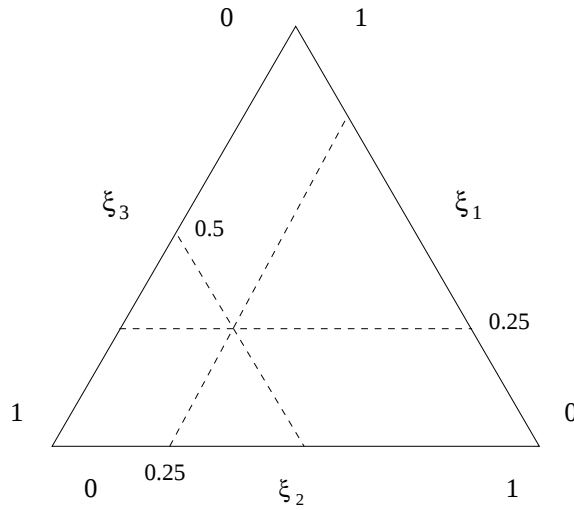


Figura 4.4: Localização do ponto $[0.25, 0.25, 0.5]$ em um simplex 3-espécies.

paralelas ao eixo ξ_1 e linhas de $\xi_3 = cte$ são paralelas ao eixo ξ_2 . O ponto onde essas linhas se interceptam corresponde ao ponto em consideração.

4.3 Análise de alguns sistemas concretos

4.3.1 Competidores Independentes

Como um exemplo simples da teoria discutida acima, consideraremos o caso de um sistema no qual todas as moléculas obedecem uma lei de crescimento linear, ou seja, a taxa de crescimento de cada espécie é proporcional a sua frequência. Neste caso temos

$$\Gamma_i = k_i x_i, \quad (4.17)$$

que corresponde a $p = 1$. Daí, as Eqs. (4.15) tomam a forma

$$\dot{\xi}_i = k_i \xi_i - \xi_i \sum_j k_j \xi_j. \quad (4.18)$$

Devemos agora determinar os pontos fixos que correspondem aos valores de ξ_i no regime estacionário ($\dot{\xi}_i = 0 \forall i$), denotados por $\bar{\xi}_i$. Esta condição resulta em

$$\bar{\xi}_i (k_i - \sum_j k_j \bar{\xi}_j) = 0. \quad (4.19)$$

Uma vez que os $\bar{\xi}_i$ são coordenadas normalizadas ($\sum_i \bar{\xi}_i = 1$), podemos facilmente verificar que uma possível solução é

$$\bar{\xi}_i = \delta_{il} \quad l = 1, \dots, n. \quad (4.20)$$

Existem, portanto, n diferentes estados de equilíbrio, cada um consistindo de um estado puro, ou seja, apenas uma única espécie está presente no regime estacionário.

Análise de estabilidade linear

Resta-nos verificar a estabilidade de cada um desses pontos fixos, onde utilizaremos o procedimento descrito no Apêndice D. Substituindo $\bar{\xi}_i$ por $\bar{\xi}_i + z_i$ nas Eqs. (4.18), onde z_i é uma pequena perturbação, e depois de algumas manipulações obtemos

$$\dot{z}_i = \left(k_i \bar{\xi}_i - \bar{\xi}_i \sum_j k_j \bar{\xi}_j \right) + z_i \left(k_i - \sum_j k_j \bar{\xi}_j \right) - \bar{\xi}_i \sum_j k_j z_j + O(z^2). \quad (4.21)$$

O primeiro termo do segundo membro é nulo conforme a Eq. (4.19). Daí,

$$\dot{z}_i = z_i \left(k_i - \sum_j k_j \bar{\xi}_j \right) - \bar{\xi}_i \sum_j k_j z_j. \quad (4.22)$$

Usando a solução $\bar{\xi}_i = \delta_{il}$, obtemos

$$\dot{z}_i = \begin{cases} -\sum_j k_j z_j & \text{se } i = l \\ z_i(k_i - k_l) & \text{se } i \neq l \end{cases}, \quad (4.23)$$

que também pode ser escrita na forma matricial

$$\dot{\mathbf{z}} = \mathbf{B}\mathbf{z}, \quad (4.24)$$

com a matrix $\mathbf{B}$ dada por

$$\mathbf{B} = \begin{bmatrix} (k_1 - k_l) & 0 & 0 & \cdots & 0 \\ 0 & (k_2 - k_l) & 0 & \cdots & 0 \\ & & \vdots & & \\ -k_1 & -k_2 & -k_3 & \cdots & -k_n \\ & & \vdots & & \\ 0 & 0 & 0 & \cdots & (k_n - k_l) \end{bmatrix}. \quad (4.25)$$

Notemos que a linha com elementos não-diagonais não-nulos corresponde a l -ésima linha. Os autovalores de B são as $n - 1$ quantidades $k_i - k_l$ para $i \neq l$, e $-k_l$. Para que o sistema apresente pelo menos um ponto de equilíbrio estável, deve existir uma situação em que todos os autovalores de B sejam negativos. Desde que as taxas de reações são necessariamente positivas, e supondo que todas elas sejam diferentes, verificamos que se k_l for a maior constante de reação, então todos os autovalores serão negativos. Se k_l for a menor taxa, então todos os autovalores (exceto $-k_l$) serão positivos. Para este e outros casos onde k_l não é a maior constante, ocorre uma mistura de autovalores positivos e negativos. Existe, portanto, apenas um único ponto de equilíbrio estável, consistindo de uma população contendo apenas uma única espécie, a que possui maior taxa de reação. A explicação para este fato é que a espécie que se reproduz mais rapidamente irá povoar todo o sistema, de tal forma que o número das outras espécies se torna insignificante perante esta.

4.3.2 Cadeias catalíticas

O exemplo dado acima não nos serviria como um modelo descrevendo o surgimento de catálise, desde que permite a sobrevivência de apenas uma única espécie, independentemente do número de espécies presentes inicialmente na população. Isto deve-se à ausência de qualquer tipo de acoplamento entre as espécies no modelo. Este é um aspecto geral de modelos que não consideram qualquer tipo de interação catalítica entre as moléculas.

Consideremos agora uma cadeia composta por n espécies moleculares onde a taxa de replicação da espécie i depende, além da sua própria concentração, da concentração da espécie $i - 1$ (exceto para espécie 1, que não é auxiliada por qualquer processo de catálise), conforme ilustrado na Figura 4.5. A equação descrevendo a variação da concentração da espécie 1 no tempo é dada por

$$\dot{x}_1 = k_1 x_1 + \phi_1, \quad (4.26)$$

onde ϕ_1 corresponde a um fluxo de diluição. Já as $n - 1$ espécies restantes podem reproduzir-se de forma autônoma ou com a ajuda de outra espécie no sistema. Temos então

$$\dot{x}_i = k_i x_i + k'_i x_i x_{i-1} + \phi_i \quad i = 2, \dots, n, \quad (4.27)$$

onde k_i e k'_i são constantes de reação. As formas para o fluxo ϕ_i dependem do tipo de vínculo que impomos ao sistema. Se adotarmos o vínculo usual de concentração constante c_0 e supusermos ainda que o termo do fluxo para cada espécie é proporcional a sua concentração, obtemos

$$\sum_i x_i = c_0 \quad e \quad \phi_i = x_i \Phi, \quad (4.28)$$

onde Φ é uma função do vetor concentração $\mathbf{x}$. O vínculo de concentração constante pode ser usado para derivar a forma de Φ . De fato, usando

$$\sum_j \dot{x}_j = 0 = \sum_{j=1}^n k_j x_j + \sum_{j=2}^n k'_j x_j x_{j-1} + \Phi \sum_{j=1}^n x_j, \quad (4.29)$$

obtemos

$$\Phi = -\frac{1}{c_0} \left[\sum_{j=1}^n k_j x_j + \sum_{j=2}^n k'_j x_j x_{j-1} \right]. \quad (4.30)$$

Assim, o sistema de equações diferenciais (4.26) - (4.27) torna-se

$$\dot{x}_1 = k_1 x_1 - \frac{x_1}{c_0} \left[\sum_{j=1}^n k_j x_j + \sum_{j=2}^n k'_j x_j x_{j-1} \right] \quad (4.31)$$

e

$$\dot{x}_i = k_i x_i + k'_i x_i x_{i-1} - \frac{x_i}{c_0} \left[\sum_{j=1}^n k_j x_j + \sum_{j=2}^n k'_j x_j x_{j-1} \right], \quad (4.32)$$

para $i = 2, \dots, n$. Como essas equações não são homogêneas, não podemos ignorar a concentração total c_0 na análise de ponto fixo.

Para este sistema existem $2n$ pontos fixos, sendo que n deles coincidem com os vértices da unidade simplex S_n e, portanto, indicam comportamento competitivo [4]. A localização

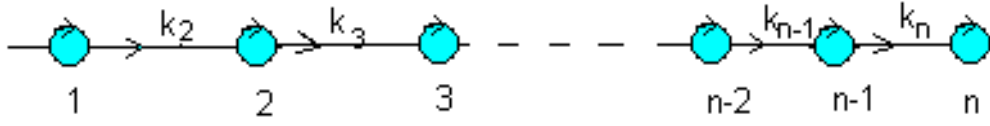


Figura 4.5: Cadeia catalítica de tamanho n .

dos outros pontos fixos no simplex depende da concentração total c_0 . O ponto fixo mais interessante para a nossa abordagem, denotado $\bar{\mathbf{x}}_{2n}$, é dado por [4, 49]

$$\bar{\mathbf{x}}_{2n} = \begin{pmatrix} \frac{k_1 - k_2}{k'_2} \\ \frac{k_1 - k_3}{k'_3} \\ \vdots \\ c_0 - \sum_{j=2}^n \frac{k_1 - k_j}{k'_j} \end{pmatrix}. \quad (4.33)$$

Se e somente se as constantes de reação satisfizerem a condição $k_1 > k_j$, $j = 2, 3, \dots, n$ e a concentração total exceder o valor crítico

$$c_{cr} = \sum_{j=2}^n \frac{k_1 - k_j}{k'_j}, \quad (4.34)$$

este ponto fixo estará localizado dentro do simplex, indicando portanto coexistência das n espécies. Caso contrário, algumas das concentrações de equilíbrio assumiriam valores negativos e, portanto, sem qualquer sentido físico.

Podemos resumir o comportamento das cadeias catalíticas da seguinte forma:

- (1) O estado estacionário estável em que as várias espécies coexistem existirá apenas quando as constantes de reação e a concentração total satisfizerem as condições:

$$k_1 > k_j \quad \forall \quad j \quad e \quad c_0 > \sum_{j=2}^n \frac{k_1 - k_j}{k'_j}. \quad (4.35)$$

- (2) Satisfeita a primeira condição, as concentrações de cada espécie só terão magnitudes comparáveis em um pequeno intervalo de valores de concentração total c_0 . Com o aumento de c_0 , o último membro da cadeia (elemento n) irá crescer continuamente e eventualmente dominará a população.

Utilizando o procedimento de análise de estabilidade linear (Ver apêndice D), podemos verificar que o ponto fixo é estável em qualquer região onde satisfaça as condições de existência acima definidas [4, 49]. Embora o resultado acima pareça a primeira vista encorajador, há certos aspectos frustrantes. Seria desejável que o ponto de equilíbrio fosse robusto a variações no meio, ou seja, mesmo no caso de variação na concentração total, o equilíbrio fosse mantido. O que não é o caso das cadeias catalíticas. Por exemplo, se c_0 cai abaixo de um certo valor crítico, o equilíbrio desaparece. Esse cenário é preocupante desde que é provável que as soluções primitivas sejam bastante diluídas. Um outro problema é que se c_0 cresce muito, apenas a n -ésima espécie amplia sua representação. De forma a tornar um sistema mais robusto e manter a diversidade do sistema, é preferível que todas as espécies aumentem em tamanho juntas.

4.3.3 Ciclos catalíticos

O conceito de hiperciclo é inspirado na propriedade cíclica das reações de síntese DNA-proteína em formas de vida mais evoluídas como os organismos modernos. O DNA contém a informação necessária para a construção das proteínas, que por sua vez são necessárias para a síntese do DNA. No ciclo catalítico consideramos todas as interações descritas para o caso da cadeia mais uma interação entre a n -ésima e a primeira espécie, formando assim um ciclo.

O sistema químico real visualizado por Eigen e Schuster [4] é da seguinte forma. Consideremos n diferentes espécies de *moléculas de informação* I_i (tal como seqüências de RNA ou DNA), cada qual contendo as instruções necessárias para a síntese de uma

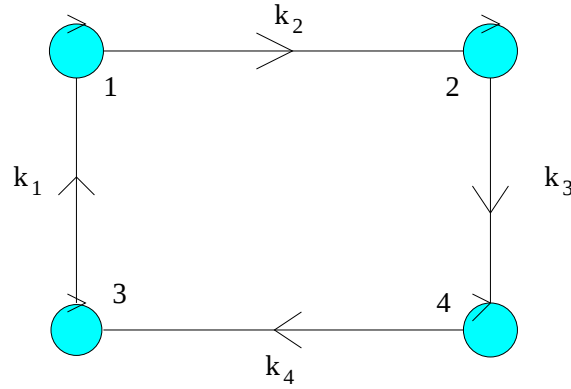


Figura 4.6: Um hiperciclo de 4 espécies.

proteína E_i . Esta proteína pode ser pensada como uma *replicase* cuja função é a de transcrever e reconhecer a próxima molécula de informação I_{i+1} na sua *replicase* E_{i+1} . Dessa forma, a taxa de variação da espécie i depende das freqüências das replicases E_{i-1} e das moléculas de informação I_i .

Existem várias maneiras pelas quais a situação descrita acima pode ser realizada, mas de forma a tornarmos a descrição mais simples, consideramos apenas termos de reações homogêneos. Como um exemplo, em um hiperciclo de 4 espécies, as interações devem ser da forma ilustrada na Fig. 4.6. Neste caso, cada espécie depende de si própria e de outra espécie para sua reprodução. O termo de reação para a espécie 1, por exemplo, é

$$\Gamma_1 = k_1 x_1 x_4, \quad (4.36)$$

pois a espécie 1 necessita de uma molécula de mesma espécie e outra da espécie 4 para a produção de uma cópia.

Um exemplo mais complicado é ilustrado na Fig. 4.7, onde temos 6 espécies, cada qual requerendo a ação de duas outras para auxiliar na sua reprodução. Neste caso, os termos de reação são todos de grau $p = 3$.

Um sistema hipercíclico geral terá n espécies no ciclo com p espécies participando em cada reação (naturalmente, em uma situação mais geral, não precisamos nos restringir

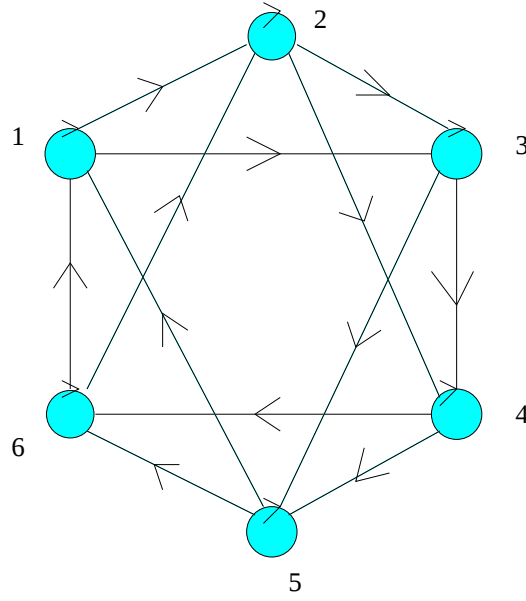


Figura 4.7: Um hiperciclo de 6 espécies.

ao mesmo número de espécies em cada reação, porém nestes casos as equações deixarão de ser homogêneas). O termo de taxas de reações em tal sistema terá a forma

$$\Gamma_i = k_i x_i x_j x_k \dots x_r \quad i = 1, \dots, n \quad (4.37)$$

com p fatores de concentração, onde os índices denotam todas as espécies participando na reação que produz uma cópia da espécie i . Se impusermos o vínculo de concentração constante, então as equações diferenciais que governam a evolução do hiperciclo são

$$\dot{x}_i = k_i x_i x_j \dots x_l - \frac{x_i}{c_0} \sum_{\{r,s,\dots,t\}} k_r x_r x_s \dots x_t. \quad (4.38)$$

A análise de ponto fixo geral é muito complexa para ser realizada analiticamente. No caso especial onde todas as constantes de reação são iguais, $k_1 = k_2 = \dots = k_n$, a análise é um tanto mais fácil [49]: para $n = 2, 3$ e 4 existe um ponto fixo estável em que todas as espécies coexistem; para $n \geq 5$, surge um ciclo limite estável em que todas as espécies participam, ou seja, todas as espécies permanecem presentes, porém oscilam em quantidade. Estudos mais aprofundados sobre o comportamento dos hiperciclos foram realizados por meio de simulações [12, 50, 51] e análises numéricas [13, 52–54].

4.4 Estudo de Propagação de Erro no Hiperciclo

Nesta seção investigamos analiticamente os estados estacionários de um sistema determinístico (população infinita, $N \rightarrow \infty$) constituído de duas unidades, o *hiperciclo* composto por n elementos auto-replicantes $I_1, \dots, I_n$ e sua *cauda de erro*, composta pelos demais tipos de moléculas [55]. Essas partes são acopladas de forma que qualquer cópia errônea dos elementos do hiperciclo pertencerá à cauda de erro. Este modelo corresponde exatamente ao limite de seqüências de tamanho infinito, $L \rightarrow \infty$. Neste limite podemos seguramente desprezar as mutações provenientes da cauda de erro para o hiperciclo, como também as mutações entre os elementos do hiperciclo.

4.4.1 O modelo

Consideramos um sistema composto de um hiperciclo de n elementos $I_1, \dots, I_n$ e sua cauda de erro, como ilustrado na Fig. 4.8. Em contraste com o então denominado hiperciclo elementar [4], assumimos que as moléculas são capazes de se auto-replicarem com taxas seletivas A_i ($i = 1, \dots, n$) e A_e para os elementos da cauda de erro. Admitimos também que a constante de interação catalítica entre os elementos I_i e I_{i-1} do hiperciclo é igual a K_i . O ingrediente chave no modelo é que em ambos os processos de crescimento da molécula de tipo i a probabilidade de sucesso é dada pelo parâmetro $Q \in [0, 1]$, de forma que uma cópia errônea, que então será atribuída a cauda de erro, é produzida com probabilidade $1 - Q$. Com isso, as concentrações x_i ($i = 1, \dots, n$) dos elementos do hiperciclo e a concentração x_e da cauda de erro evoluem no tempo de acordo com as equações cinéticas

$$\dot{x}_i = x_i(A_i Q + K_i x_{i-1} Q - \Phi) \quad i = 1, \dots, n \quad (4.39)$$

e

$$\dot{x}_e = x_e(A_e - \Phi) + (1 - Q) \sum_{i=1}^n x_i(A_i + K_i x_{i-1}), \quad (4.40)$$

onde $x_0 \equiv x_n$ e

$$\Phi = \sum_{i=1}^n x_i (A_i + K_i x_{i-1}) + A_e x_e \quad (4.41)$$

é um fluxo de diluição de forma a manter a concentração constante, isto é, $\sum_{i=1}^n \dot{x}_i + \dot{x}_e = 0$.

Assumimos que

$$\sum_{i=1}^n x_i + x_e = 1. \quad (4.42)$$

Esta formulação é equivalente a considerar polinucleotídeos de tamanho $L \rightarrow \infty$ cuja probabilidade de replicação exata por nucleotídeo tende a 1 e a probabilidade de replicação exata por genoma é finita, ou seja, $q^L \rightarrow Q$. Neste limite as mutações provenientes da cauda de erro para os elementos do hiperciclo, como também as mutações entre os elementos do hiperciclo, podem ser seguramente desprezadas. Portanto, as mutações podem apenas aumentar a concentração dos elementos compondo a cauda de erro. A vantagem de trabalhar neste limite é que a transição do limiar de erro pode ser precisamente localizada por meio da determinação do valor de Q em que a concentração de uma molécula relevante (seqüência mestra) se anula. Para valores finitos de L , como também para tamanhos finitos de população, a caracterização desta transição é mais complicada, sendo necessário o uso da técnica de escalonamento de tamanho finito [28].

Neste trabalho consideramos o relevo replicativo de pico único, em que atribuímos valor seletivo $A_1 = a > 1$ à então denominada seqüência mestra I_1 , e $A_i = A_e = 1$ para os demais elementos do hiperciclo, como também para todos os elementos compondo a cauda de erro. Por simplicidade adotamos $K_i = K \forall i$. A motivação desta escolha particular de parâmetros é a observação de que a emergência do hiperciclo requer tanto uma coexistência espacial como temporal das seqüências compondo a rede, e isto pode ser obtido por meio de uma distribuição de quase-espécies, que garante a coexistência da seqüência e de seus mutantes mais próximos, apesar do caráter puramente competitivo do modelo de quase-espécies [51]. Uma vez que a coexistência é estabelecida, o surgimento de acopla-

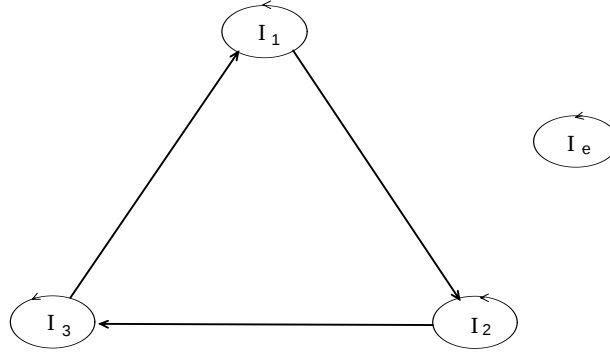


Figura 4.8: Sistema composto de um hiperciclo de tamanho $n = 3$ e sua cauda de erro I_e . As flechas finas representam as reações de auto-replicação não catalizadas e as flechas grossas indicam a auto-replicação cataliticamente assistida pela seqüência vizinha.

mentos catalíticos entre as seqüências não é um evento muito improvável. Naturalmente, tão logo esses acoplamentos cooperativos tornem-se suficientemente fortes, de forma a balancear a competição imposta pelo vínculo de concentração constante, os mutantes irão certamente partir da seqüência mestra devido à implacável pressão das mutações, de forma que nenhum traço permanecerá da distribuição original da quase-espécies.

4.4.2 Análise de pontos fixos

Vamos agora fazer algumas considerações que serão importantes no decorrer da análise do modelo.

Vamos fazer uma distinção entre seqüências sobreviventes, $x_i > 0$, e seqüências extintas, $x_i = 0$. Um sobrevivente I_j é dito isolado se $x_{j-1} = x_{j+1} = 0$. Portanto,

$$\dot{x}_j = x_j(Q - \Phi) \quad j > 1 \quad (4.43)$$

e

$$\dot{x}_e = x_e(1 - \Phi) + (1 - Q) \left[x_1(a - 1) + 1 + K \sum_{i \neq j, j+1} x_i x_{i-1} \right]. \quad (4.44)$$

No regime estacionário, a Eq. (4.43) nos dá $\Phi = Q$ que, para $Q \leq 1$, é incompatível com Eq. (4.44), desde que o termo entre os colchetes nesta equação é positivo. Portanto, todos os sobreviventes isolados, com a exceção da seqüência mestra, são instáveis contra a cauda de erro. Consideremos agora a seguinte cadeia de seqüências sobreviventes:

$$\dot{x}_i = x_i(Q - \Phi) \quad (4.45)$$

$$\dot{x}_{i+1} = x_{i+1}(Q + KQx_i - \Phi) \quad (4.46)$$

$$\dot{x}_{i+2} = x_{i+2}(Q + KQx_{i+1} - \Phi) \quad (4.47)$$

$$\vdots \quad (4.48)$$

$$\dot{x}_k = x_k(Q + KQx_{k-1} - \Phi) \quad (4.49)$$

que não contém x_1 . Novamente, no estado de regime estacionário a primeira equação nos dá $\Phi = Q$, implicando $KQx_i = 0$, ou seja, $x_i = 0$. Desta forma, não há ponto fixo correspondente a tal cadeia. Qualquer cadeia de sobreviventes deve portanto começar com as seqüências n ou 1 . No primeiro caso obtemos $\Phi = Q$ de $\dot{x}_n = 0$ e $aQ + KQx_n = Q$ de $\dot{x}_1 = 0$, resultando $x_n = (1 - a)/K \leq 0$, que descarta esta possibilidade. Os pontos de equilíbrio de interesse para nosso estudo são desta forma o equilíbrio em que todas as seqüências sobrevivem, ou um ponto fixo que corresponde a uma cadeia de sobreviventes começando com I_1 .

De acordo com isto, definimos um estado de coexistência de m -elementos, ℓ_m , como o vetor de n -componentes $\mathbf{x} = (x_1, x_2, \dots, x_n)$ em que os primeiros m -elementos são estritamente positivos e o restante iguais a zero. Claramente, dado o vetor de seqüências $\mathbf{x}$, a concentração da cauda de erro x_e é determinada pelo vínculo (4.42). Daqui em diante, resolveremos analiticamente as equações cinéticas no regime de estado estacionário $\dot{x}_i = 0 \forall i$.

O ponto fixo mais simples consiste do estado ℓ_0 ($m=0$), que corresponde à solução $x_1 = \dots = x_n = 0$ e $x_e = 1$, válido para todos os valores de parâmetro.

No caso de cadeias, $0 < m < n$, as soluções de estado estacionário das Eqns. (4.39) e (4.40) são bastante simples. De fato, desde que $x_n = 0$ por definição, obtemos que $\Phi = aQ$ de $\dot{x}_1 = 0$, que resulta em

$$x_1 = x_2 = \dots = x_{m-1} = \frac{a-1}{K}. \quad (4.50)$$

Utilizando este resultado na Eq.(4.41), obtemos

$$x_m = \frac{Qa-1}{a-1} - (m-1)\frac{a-1}{K}. \quad (4.51)$$

Entretanto, desde que $x_i \in (0, 1) \forall i$, esta solução é fisicamente significativa apenas na região $K > a-1$ e $Q > Q_m$, onde

$$Q_m = \frac{1}{a} + \frac{1}{Ka}(m-1)(a-1)^2. \quad (4.52)$$

Notemos que o estado ℓ_1 (quase-espécies) é obtido fazendo $m = 1$ na Eq. (4.51) e sua região de existência é simplesmente $Q > 1/a$, desde que a outra condição, $K > a-1$, é derivada considerando as outras seqüências na cadeia. De fato, este resultado bastante simples quantifica a noção de que os acoplamentos cooperativos devem alcançar um valor mínimo de forma a contra-balancear a competição entre as seqüências.

A análise dos hiperciclos, isto é, $m = n$, é um tanto mais complicada. Da condição $\dot{x}_2$ obtemos $\Phi - Q = KQx_1$ que, inserida nas equações $\dot{x}_1 = \dots = \dot{x}_n = 0$, resulta em $x_1 = x_2 = \dots = x_{n-1}$ e

$$x_n = x_1 - \frac{a-1}{K}. \quad (4.53)$$

Finalmente utilizando estes resultados na Eq. (4.41), encontramos que x_1 é dado pelas raízes da equação quadrática

$$nKx_1^2 - (KQ + a-1)x_1 + 1 - Q = 0. \quad (4.54)$$

Para $K < (a-1)^2/4n$ esta equação tem raízes reais para todo $Q \geq 0$, caso contrário ela possui raízes reais para $Q \geq Q_h$, onde Q_h é a única raiz positiva da equação

$$K^2 Q_h^2 + 2K(a-1+2n)Q_h + (a-1)^2 - 4nK = 0. \quad (4.55)$$

Em particular para grandes valores de K encontramos $Q_h \approx 2\sqrt{n/K}$. Além do mais, pode ser visto facilmente das Eqns. (4.53) e (4.54) que x_n se anula em $Q = Q_n$ com Q_n dado por Eq. (4.52). Para entender o papel de Q_h e Q_n (notemos que $Q_n \geq Q_h$) em delimitar a região de existência do estado ℓ_n , devemos nos entreter no comportamento das duas raízes reais da Eq. (4.54). Sejam elas x_1^+ e x_1^- com $x_1^+ \geq x_1^-$ que, de acordo com Eq. (4.53) corresponde a x_n^+ e x_n^- , respectivamente. Naturalmente, essas raízes tornam-se idênticas em $Q = Q_h$ e então as duas soluções para x_n se anularão simultaneamente apenas para $K = K_h$, onde Q_h iguala Q_n . Explicitamente, obtemos

$$K_h = (a-1)[n(a+1) - 1], \quad (4.56)$$

inserindo Eq. (4.52) na Eq. (4.55). Embora ambas as raízes x_1^+ e x_1^- estejam no simplex $(0,1)$, isto não é verdade para x_n^+ e x_n^- . Em particular, quando $K < K_h$ ambas as concentrações são negativas dentro do intervalo $Q_h \leq Q \leq Q_n$. Entretanto, enquanto x_n^+ torna-se positivo para $Q > Q_n$ (ela se anula em Q_n), x_n^- permanece sempre negativa. Desde que $K_h > (a-1)^2/4n$, a mesma conclusão é válida para o intervalo $K < (a-1)^2/4n$, se definirmos $Q_h = 0$ nesta região. Esta situação é inversa para $K > K_h$: ambas as concentrações são positivas no intervalo $Q_h \leq Q \leq Q_n$, mas neste caso é x_n^- que se anula em Q_n e torna-se negativo para $Q \leq Q_n$, enquanto x_n^+ permanece sempre positivo. Apesar da pequena região no espaço de parâmetros onde a raiz x_1^- resulta em concentrações no interior do simplex, a análise de estabilidade linear discutida posteriormente indica que esta solução é sempre instável, de forma que precisamos apenas considerar a raiz x_1^+ . Desta forma o intervalo de existência do ponto fixo para hiperciclo $m = n$ é $Q \geq Q_n$ se $K \leq K_h$, e $Q \geq Q_h$ se $K \geq K_h$.

4.4.3 Análise de estabilidade

Para realizarmos a análise de estabilidade linear dos pontos fixos obtidos na seção anterior, é conveniente reescrevermos as equações cinéticas (4.39) e (4.40) como

$$\dot{x}_i = x_i F_i(\mathbf{x}) \quad i = 1, \dots, n, \quad (4.57)$$

onde

$$F_i(\mathbf{x}) = A_i Q + K Q x_{i-1} - A_e - \sum_j x_j (A_j - A_e + K x_{j-1}), \quad (4.58)$$

e de forma a eliminarmos x_e na equação acima utilizamos o vínculo (4.42). A estabilidade de um ponto fixo é assegurada dado que as partes reais de todos os autovalores do Jacobiano $\mathcal{J}$ $n \times n$ sejam negativas (Ver apêndice D). Em nosso caso os elementos do Jacobiano são dados por

$$J_{ij} = \delta_{ij} F_i + x_i \frac{\partial F_i}{\partial x_j} \quad i, j = 1, \dots, n. \quad (4.59)$$

O cálculo dos autovalores é simples apenas para o estado ℓ_0 , pois neste caso o Jacobiano é diagonal com elementos $J_{11} = aQ - 1$ e $J_{ii} = Q - 1$, $i \geq 1$. Portanto, os estados estacionários tornam-se instáveis para $Q > 1/a$, que coincide com o valor mínimo na precisão de replicação necessária para a existência do estado ℓ_1 . Entretanto, para um estado geral ℓ_m temos que recorrer a estimativas numéricas dos autovalores do Jacobiano.

Felizmente no caso de cadeias, $0 < m < n$, há uma alternativa para determinarmos a estabilidade dos pontos fixos, como aquela realizada para o caso ℓ_0 , que torna-se instável devido a emergência do estado ℓ_1 . De fato, pode ser facilmente verificado que qualquer perturbação do ponto fixo de coexistência de m -espécies que torne a concentração x_{m+1} não-nula será amplificada, se $A_{m+1}Q + KQx_m - \Phi$ é positivo. Para $m > 0$, utilizamos $\Phi = aQ$ e $A_{m+1} = 1$ juntamente com o valor de x_m , dado pela Eq. (4.51), de forma a obtermos a seguinte condição para a estabilidade do estado ℓ_m

$$Q \leq Q_{m+1} \quad m > 0, \quad (4.60)$$

com Q_m descrito pela Eq. (4.52). Portanto, o valor máximo de Q permitido para a estabilidade do estado ℓ_m coincide com o valor mínimo necessário para a existência do estado ℓ_{m+1} . Interessantemente, embora para $m = 0$ tenhamos $\Phi = 1$, $A_1 = a$ e $x_0 = 0$, a condição (4.60) também é verdadeira para este caso.

Neste ponto duas considerações são importantes. Primeiramente, o argumento levando à condição de estabilidade (4.60) é falho se o estado ℓ_{m+1} é instável. Portanto, devemos garantir via cálculo numérico dos autovalores do Jacobiano que o estado ℓ_l é estável antes de usar esta condição para estudar a estabilidade de cadeias com $m \leq l$. Em particular, realizamos análises numéricas para a solução do hiperciclo $l = n$ e encontramos os seguintes resultados:

- (i) Para $n \leq 4$, o hiperciclo é sempre estável;
- (ii) Para $n = 5$, o hiperciclo é estável em um pequeno intervalo de Q acima de Q_5 e;
- (iii) Para $n \geq 6$, o hiperciclo é sempre instável.

Em segundo lugar, a derivação da condição de estabilidade (4.60) é baseada na análise de um único autovalor do Jacobiano e portanto não nos dá uma condição suficiente para a estabilidade dos pontos fixos. Contudo, verificamos por meio dos dados numéricos de todos os n autovalores que, dado que o estado ℓ_{m+1} é estável, o autovalor associado às flutuações levando a um aumento do tamanho da cadeia é o primeiro a se tornar positivo.

4.4.4 Discussão

Combinando os resultados da existência e estabilidade derivados nas seções anteriores, podemos então construir os diagramas de fases no plano (K, Q) para valores fixos de a e n . Em particular, para $n \leq 4$ o estado ℓ_m é estável dentro do intervalo

$$Q_m \leq Q \leq Q_{m+1} \quad m \leq n, \quad (4.61)$$

com Q_m dado por Eq. (4.52), desde que $K > a - 1$. Interessantemente, para valor fixo de Q , Eq. (4.61) nos mostra que o incremento δK na constante catalítica necessária para incorporar uma nova seqüência na cadeia é

$$\delta K = \frac{(a-1)^2}{aQ-1}, \quad (4.62)$$

independentemente do número de elementos na cadeia.

O caso $K < a-1$, para o qual nenhuma cadeia com $m > 1$ é permitida, não requer nenhuma consideração especial. De fato, encontramos que os únicos estados estáveis são ℓ_0 ($Q \leq 1/a$) e ℓ_1 ($1/a \leq Q \leq 1$). Entretanto, desde que $Q_2 \leq 1$ apenas para $K \geq a-1$, este resultado é consistente com Eq. (4.61).

O estado ℓ_n (ou seja, solução de hiperciclo) é estável para $Q \geq Q_n$ se $K \leq K_h$ e para $Q \geq Q_h$ caso contrário, onde K_h e Q_h são descritas pelas Eqs. (4.56) e (4.55), respectivamente. Definimos o limiar de erro do hiperciclo como o valor de replicação exata Q que delimita a região de estabilidade de ℓ_n . O diagrama de fases para $a = 10$ e $n = 4$ exibido na Fig. 4.9 ilustra o papel principal de K_h na organização hipercíclica: apenas para $K > K_h$ o hiperciclo torna-se mais estável que uma cadeia de mesmo tamanho. Uma outra quantidade importante é o valor de K , representado por K_c , em que Q_h é igual a $1/a$, a precisão mínima de replicação da quase-espécie, que é igual a

$$K_c = a(a-1) \left[2n-1 + \sqrt{n(n-1)} \right]. \quad (4.63)$$

Além deste valor, o limiar de erro do hiperciclo Q_h é menor que o observado no modelo de quase-espécies. Além do mais, como mencionado anteriormente, para grandes valores de K , ele vai a zero com $1/\sqrt{K}$.

Um aspecto muito frustrante de K_h e K_c é que ambos são da ordem de a^2 , indicando que a produtividade de auto-replicação cataliticamente assistida é muito maior que a de auto-replicação não-catalizada. Enquanto isto é obviamente verdadeiro para catálise

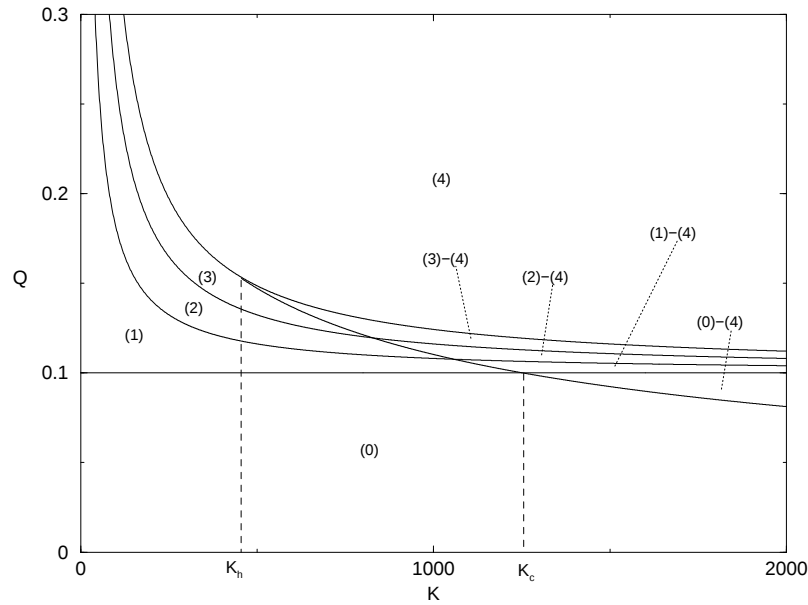


Figura 4.9: Diagrama de fases no espaço (K, Q) para $n = 4$ e $a = 10$ exibindo as regiões de estabilidade dos diversos estados de coexistência. Os números entre parênteses indicam o número de espécies coexistentes. Regiões de biestabilidade aparecem para $K \geq K_h$. As linhas finas são (de baixo para cima) Q_1 , Q_2 , Q_3 e Q_4 . A linha cheia é Q_h .

bioquímica, é difícil justificar a existência de catalizadores tão eficientes em condições pré-bióticas. Por outro lado, podemos tomar um ponto de vista diferente e mais otimista e, argumentar que catalizadores bioquímicos modernos (enzimas) são tão eficientes por causa de seus precursores que tinham que satisfazer a rigorosa condição de superar K_h .

Na Fig. 4.10 apresentamos o diagrama para $n = 5$. A principal diferença com relação à figura anterior é que o estado ℓ_5 é estável apenas dentro de uma pequena região entre Q_5 e a linha tracejada, obtida por meio de cálculo numérico dos autovalores do Jacobiano. Como estas curvas se interceptam em algum $K \leq K_h$, o hiperciclo de 5-membros não é muito interessante, desde que possuem as mesmas características de uma cadeia de tamanho $m = 5$. Para confirmar estes resultados, realizamos integração numérica das equações cinéticas usando o método de Runge-Kutta de nona-ordem [46]. Os resultados são exibidos na Fig. 4.11, onde ilustramos a evolução temporal das concentrações x_i ($i = 1, \dots, 5$) dentro e fora da região de estabilidade.

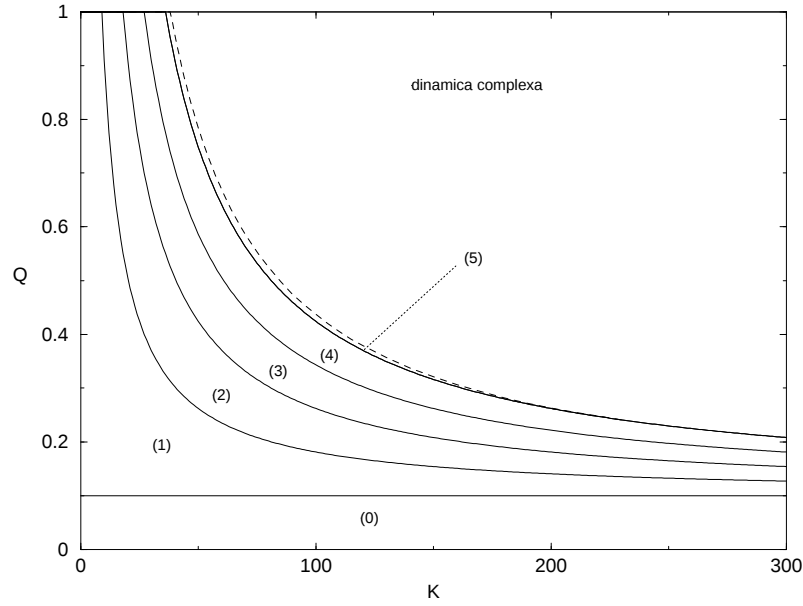


Figura 4.10: O mesmo que Fig. 4.9, porém $n = 5$. Não há pontos fixos estáveis acima da linha tracejada. As curvas sólidas são (de baixo para cima) Q_1, Q_2, Q_3, Q_4 e Q_5 .

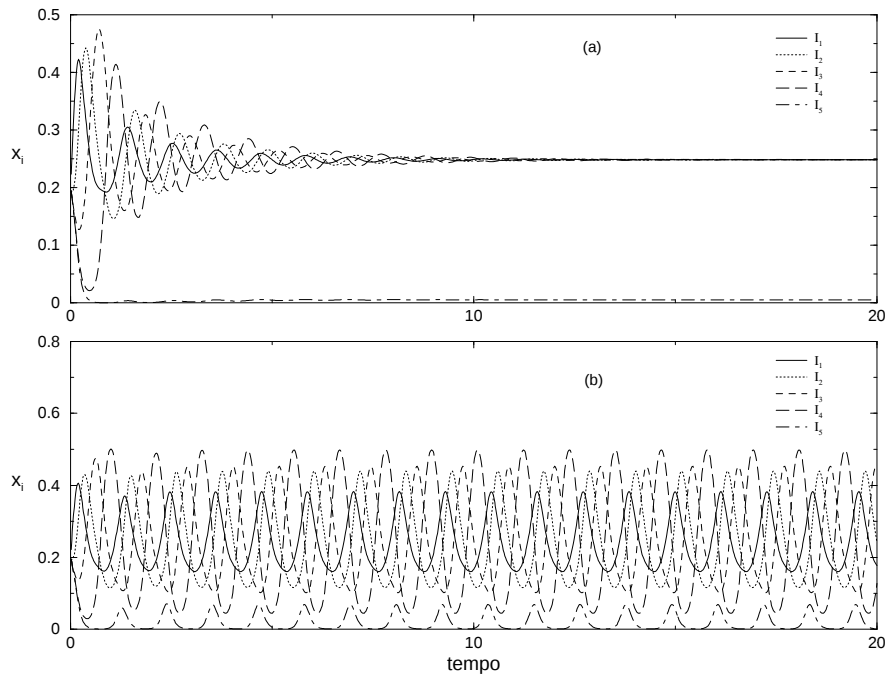


Figura 4.11: Evolução temporal das cinco concentrações dos elementos compondo o hiperciclo de tamanho $n = 5$ para $a = 10$, $Q = 1$, e (a) $K = 37$ (dentro da região de estabilidade) e (b) $K = 40$ (fora da região de estabilidade). O estado inicial é $x_i = 0.2 \forall i$.

Embora o padrão de comportamento na região de instabilidade seja aparentemente periódico, não exploramos completamente o espaço de parâmetros para descartar a existência de um comportamento caótico. Assim sendo, utilizamos o termo dinâmica complexa para caracterizar esta região na Fig. 4.10. Observamos que o diagrama de fases exibido nesta figura descreve também as regiões de estabilidade de hiperciclos e cadeias de tamanho $n \geq 5$, desde que os estados ℓ_m com $m > 5$ são sempre instáveis.

Um caso limitante interessante que merece atenção especial é o hiperciclo simétrico ($a = 1$). De acordo com o argumento discutido no início da seção anterior, os únicos pontos fixos neste caso são os estados ℓ_0 e ℓ_n (hiperciclo), ou seja, cadeias não são permitidas. Além do mais, Eq. (4.53) nos dá $x_1 = x_2 = \dots = x_n$ onde x_1 é descrito pela Eq. (4.54) com a substituído por 1. A análise das raízes da equação quadrática e os dados numéricos dos autovalores do Jacobiano indicam que o hiperciclo simétrico é estável para

$$K > \frac{4n}{Q^2}(1 - Q), \quad (4.64)$$

quando $n \leq 4$. A região de estabilidade observada na Fig. 4.10 para o estado ℓ_5 não aparece no caso simétrico $a = 1$, e portanto isto deve ser uma consequência da assimetria nos valores de produtividade das reações de auto-replicação não-catalizadas. Observamos que, diferentemente do caso assimétrico ($a > 1$), o estado ℓ_0 é sempre estável.

O hiperciclo elementar

Finalizando as discussões, apresentamos alguns resultados para o hiperciclo elementar ($A_i = 0$, $i = 1, \dots, n$) acoplado à cauda de erro por meio de mutações durante o processo de auto-replicação cataliticamente assistida. Inserindo estes parâmetros na Eq. (4.58) e fazendo $F_i = 0 \forall i$, obtemos $x_1 = x_2 = \dots = x_n$ com x_1 dado pela maior raiz da equação quadrática

$$Kn x_1^2 - (n + KQ)x_1 + 1 = 0, \quad (4.65)$$

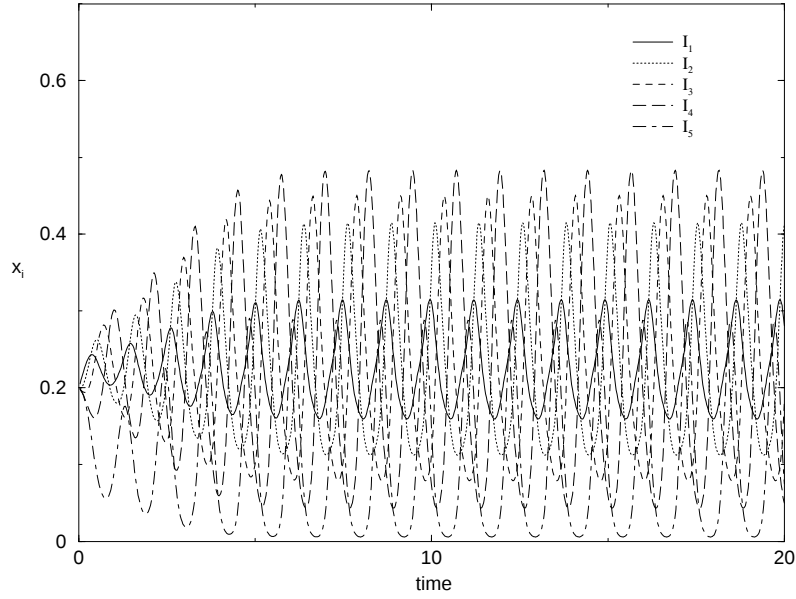


Figura 4.12: Evolução temporal das cinco concentrações dos elementos compondo uma cadeia de tamanho $n = 5$. Os parâmetros e o estado inicial inicial são os mesmos utilizados para 4.11(b).

pois verificamos por meio de análise numérica que a menor raiz é sempre instável. Como no caso simétrico discutido acima, para $n \leq 4$ a condição de estabilidade coincide com a condição para x_1 real,

$$Q \geq 2\sqrt{\frac{n}{K}} - \frac{n}{K}. \quad (4.66)$$

Assim, o termo do lado direito desta desigualdade nos dá o limiar de erro do hiperciclo elementar.

Capítulo 5

Compartimentalização em Evolução Pré-biótica

Nesta capítulo, discutiremos modelos alternativos ao hiperciclo que também têm por finalidade a coexistência de elementos estruturalmente distintos numa população de macromoléculas [56]. Esse estudo é aqui abordado no contexto da *Teoria de Seleção de Grupo*.

Evolução e manutenção de altruísmo tornou-se um tópico de interesse no campo de evolução pré-biótica desde que Maynard Smith [15] considerou a atividade catalítica desempenhada por uma macromolécula numa rede de reações catalíticas, como o hiperciclo, como de fato um comportamento altruísta. Definimos por comportamento altruísta aquele que é de alguma forma prejudicial ao próprio elemento que o desempenha, mas que confere uma vantagem seletiva ao grupo do qual é parte [57].

Na modelagem tradicional de seleção de grupo, baseada nos modelos de *ilhas* de Wright [58, 59], assumimos que a população é dividida em sub-populações reprodutíveis isoladas, ou *demes*. A estabilidade dos elementos altruístas é assegurada por meio da existência de um mecanismo de extinção que, ocorre a uma taxa dependente da composição das *demes*. Naturalmente, tais extinções favorecem a ocorrência de indivíduos que reduzem a probabilidade de extinção das *demes* das quais fazem parte, no caso os elementos altruístas [60–62]. Uma formulação mais moderna de seleção de grupo, elaborada

por Wilson [22], considera as *demes* como grupos característicos, em que na verdade as interações ecológicas, bioquímicas ou sociais ocorrem, mas os indivíduos são permitidos acessar e competir pelos recursos totais disponíveis no meio. Claramente, nesta formulação a noção de grupo é um tanto obscura, desde que há um estágio do ciclo de vida dos indivíduos em que estes abandonam suas *demes* para interagir com toda a população.

Na verdade, não é difícil visualizar sistemas físicos descritos pelos grupos característicos de Wilson ou modelo de *demes* estruturados. Como exemplo, alguns aspectos básicos da dinâmica de seleção viral pode ser modelada visualizando as células como *demes* [63, 64]. Neste caso, supomos que apenas N vírus infectam a célula; entretanto dentro da célula, os vírus sofrem um crescimento exponencial até o seu rompimento e a conseqüente liberação de todos os vírus, que novamente infectarão outras células, e assim por diante. Como apenas N vírus podem infectar cada célula, há uma competição efetiva entre todos os indivíduos na população. Essa formulação também é bem adequada para descrever evolução *in vitro* de vírus [64]. Uma outra aplicação interessante do formalismo desenvolvido por Wilson é a evolução de produção de enzimas no contexto pré-biótico [17, 65]. A motivação para esse tipo de estudo advém do provável cenário para a sopa primitiva, onde certamente a distribuição dessas macromoléculas não era homogênea. Fendas em rochas e partículas de poeira suspensas [23] seriam os habitats locais (*demes*) dentro dos quais essas moléculas seriam distribuídas e sofreriam crescimento exponencial. Em adição, Woese [24] e Towe [25] sugeriram um cenário atmosférico para a origem da vida em gotas de água suspensas que, seriam dispersadas por meio da ação do vento.

A formulação matemática do modelo de *demes* estruturados é centrado no conceito de frequências médias locais de altruístas, que são definidas como as frequências percebidas pelo altruísta e não-altruísta médio na população [22]. Essas quantidades diferem da frequência global de altruístas devido à variância não-nula da distribuição da

composição das *demes*, ou seja, a população não é homogênea. Em particular, a estabilidade dos altruístas e não-altruístas é assegurada supondo que os valores seletivos de altruístas e não-altruístas sejam proporcionais a suas frequências locais.

5.1 O Modelo

A população é composta de um número infinito de *demes*, cada qual composta por N indivíduos haplóides, que se reproduzem assexuadamente. Os indivíduos compondo esta população são de dois tipos possíveis, A ou B , se têm comportamento altruísta ou não-altruísta, respectivamente. Por definição, indivíduos altruístas aumentam o valor seletivo de todos os indivíduos no grupo do qual fazem parte, o que acarreta um custo para este indivíduo, que é traduzido numa redução de seu valor seletivo individual. Desta forma, o ingrediente chave de qualquer modelo de seleção de grupo é que o valor seletivo dos indivíduos dependem da composição da *deme*, que são classificadas de acordo com o número de altruístas que possuem. Há então $N + 1$ diferentes tipos de *demes*, rotuladas pelos números inteiros $i = 0, 1, \dots, N$. Portanto, um indivíduo altruísta pertencendo a um *deme* de tipo i tem valor seletivo $F_A(i)$, enquanto que um indivíduo não-altruísta tem valor seletivo $F_B(i)$, onde $F_B(i) \geq F_A(i)$. Em nosso modelo, consideramos que os indivíduos mutam para a outra classe com probabilidade $u \in [0, 1/2]$.

De forma a derivarmos uma equação de recorrência para a frequência de altruístas p_t na geração t , é mais conveniente introduzirmos a frequência de demes com $i = 0, \dots, N$ altruístas na geração t , denotada por $Y_t(i)$. Supondo que não há sobreposição de gerações, o número médio de altruístas $\mathcal{N}_A$ e não-altruístas $\mathcal{N}_B$ gerados durante o estágio de crescimento ilimitado dentro das *demes* são

$$\mathcal{N}_A = \sum_{i=0}^N [(1-u) i F_A(i) + u (N-i) F_B(i)] Y_t(i) \quad (5.1)$$

e

$$\mathcal{N}_B = \sum_{i=0}^N [(1-u)(N-i)F_B(i) + uiF_A(i)] Y_t(i), \quad (5.2)$$

respectivamente. Portanto, a frequência global de altruístas na população na geração $t+1$ é dada por

$$p_{t+1} = \frac{\mathcal{N}_A}{\mathcal{N}_A + \mathcal{N}_B}, \quad (5.3)$$

onde

$$w_t = \sum_{i=0}^N [iF_A(i) + (N-i)F_B(i)] Y_t(i) \quad (5.4)$$

é o valor seletivo médio da população. O próximo passo na modelagem corresponde a distribuição dos indivíduos nos infinitos *demes*, onde cada qual contém exatamente N indivíduos. Supomos que os indivíduos são alocados em cada *deme* de forma aleatória, o que nos leva a uma distribuição binomial

$$Y_{t+1}(i) = \binom{N}{i} (p_{t+1})^i (1-p_{t+1})^{N-i}, \quad (5.5)$$

que junto com as Equações (5.3) e (5.4) permitem uma completa descrição do ciclo de vida dos indivíduos.

Vamos agora brevemente introduzir os aspectos básicos do formalismo original do modelo de *demes* estruturados elaborado por Wilson [22]. As distribuições de probabilidade condicional de replicadores tipo A dado tipo $l = A, B$ na geração t são definidos por

$$\mathcal{P}_t(i|A) = \frac{iY_t(i)}{\sum_{i=0}^N iY_t(i)}, \quad (5.6)$$

$$\mathcal{P}_t(i|B) = \frac{(N-i)Y_t(i)}{\sum_{i=0}^N (N-i)Y_t(i)}, \quad (5.7)$$

que deve ser interpretado da seguinte maneira: considerando um tipo particular de replicador $l = A, B$ então a probabilidade $\mathcal{P}_t(i|l)$ é a probabilidade desse replicador pertencer a uma *deme* contendo i indivíduos altruístas. Assim, a frequência local média de altruístas

experimentadas pelos altruístas é dada por

$$f_A(t) = \frac{1}{N} \sum_{i=0}^N i \mathcal{P}_t(i|A) = p_t + \frac{\sigma_t^2}{N^2 p_t}, \quad (5.8)$$

onde $\sigma_t^2 = \sum_i i^2 Y_t(i) - [\sum_i i Y_t(i)]^2$ é a variância da distribuição de *demes* e $p_t = \sum_i i Y_t(i)/N$ é a frequência global de altruístas na população. Igualmente, a frequência local média de altruístas sentida por elementos não-altruístas é igual a

$$f_B(t) = \frac{1}{N} \sum_{i=0}^N i \mathcal{P}_t(i|B) = p_t - \frac{\sigma_t^2}{N^2(1-p_t)}. \quad (5.9)$$

No caso onde as *demes* são formadas de forma aleatória, obedecendo uma binomial, temos que $\sigma_t^2 = N p_t(1-p_t)$ de forma que para N grande as frequências locais tendem a 1. Desde que, $f_A \geq p_t \geq f_B$, o aspecto principal quando introduzimos frequências locais é mostrar que uma população estruturada em grupos de composição distinta pode simultaneamente melhorar os efeitos da presença de altruístas neles próprios e reduzir os efeitos benéficos nos elementos não-altruístas. Naturalmente, a suposição de que a distribuição da composição das *demes* $Y_t(i)$ afeta a dinâmica apenas por meio das frequências locais $f_l(t)$, $l = A, B$ é bastante restritiva, limitando, por exemplo, as escolhas para a dependência dos valores seletivos dos indivíduos na composição da *deme*.

5.2 Evolução de produção de enzimas

De acordo com o cenário proposto por Michod [17], consideramos dois tipos de replicadores, A e B , onde assumimos que o replicador A é capaz de produzir um catalizador (enzima) que, entretanto, pode catalizar a replicação de ambos os tipos de replicadores, mas com eficiências distintas. Desde que o replicador A , que produz o catalizador, deve ter algum custo em sua taxa de auto-replicação, enquanto B atinge todos os benefícios da catálise sem ter qualquer custo para a sua reprodução, temos então uma situação típica de comportamento altruísta. O custo associado com esse comportamento altruísta é modelado atribuindo-se uma taxa de auto-replicação $1 - r$, com $r \in [0, 1]$, para replicador

A e taxa de auto-replicação 1 para o elemento B . Além do mais, a taxa de replicação catalizada é proporcional à concentração de enzimas na *deme*, que por sua vez é proporcional à concentração de replicadores de tipo A naquela *deme*. Portanto, supondo que a auto-replicação e a replicação catalizada pela enzima são processos distintos, o valor seletivo de um replicador $l = A, B$ pertencente a um *deme* com i elementos altruístas é dado por

$$F_l(i) = 1 - \alpha_l r + K_l \frac{i}{N} \quad i = \alpha_l, \alpha_l + 1, \dots, N - 1 + \alpha_l, \quad (5.10)$$

onde $\alpha_l = 1$ se $l = A$ e, 0 se $l = B$. Aqui os parâmetros k_l representam o efeito benéfico da replicação mediada pela enzima. Em particular, $k_B = 0$ implica que a enzima é específica para o replicador que a produz, como no caso do hiperciclo composto por um único elemento. Entretanto, seria mais plausível assumir que as enzimas primordiais eram algum tipo de catalizador geral que, facilitava a replicação de um grande espectro de replicadores. Assim, assumimos que $k_A \geq k_B$. Em seu artigo, Michod considera apenas o caso especial onde $k_A = k_B$ e $u = 0$ [17].

A equação de recorrência para a concentração de altruístas é então dada por

$$p_{t+1} = \frac{1}{w_{tot}} \left[\sum_{i=0}^N i F_A(i) (1 - u) + (N - i) F_B(i) u \right] Y_t(i), \quad (5.11)$$

com F_A e F_B descritos por Eq. (5.10). Substituindo Eq. (5.10) na equação acima, obtemos

$$\begin{aligned} p_{t+1} &= \frac{1}{w_{tot}} \sum_{i=0}^N \left[\left((1 - r) + \frac{k_A}{N} i \right) i (1 - u) + \left(1 + \frac{k_B}{N} i \right) (N - i) u \right] \binom{N}{i} p_t^i (1 - p_t)^{N-i} \\ &= \frac{1}{w_{tot}} \left[(1 - r)(1 - u) N p_t + \frac{k_A}{N} (1 - u) (N^2 p_t^2 + N p_t (1 - p_t)) \right. \\ &\quad \left. + u \left(N - N p_t + k_B N p_t - \frac{k_B}{N} (N^2 p_t^2 + N p_t (1 - p_t)) \right) \right] \\ &= \frac{1}{w_{tot}} \left[(1 - u) p_t ((1 - r) N + k_A N p_t + k_A (1 - p_t)) \right. \\ &\quad \left. + u (N - N p_t + k_B N p_t - k_B N p_t^2 - k_B p_t (1 - p_t)) \right] \\ &= \frac{1}{w_{tot}} \left[(1 - 2u) [(1 - r) N p_t + k_A N p_t^2 + k_A p_t (1 - p_t)] \right. \end{aligned}$$

$$+u \left[N + Np_t(k_B - r) + N(k_A - k_B) \left(\frac{1}{N}p_t(1 - p_t) + p_t^2 \right) \right], \quad (5.12)$$

onde

$$\begin{aligned} w_{tot} &= \sum_{i=0}^N \left[\left((1-r) + k_A \frac{i}{N} \right) i + \left(1 + \frac{k_B}{N} i \right) (N-i) \right] \binom{N}{i} p_t^i (1-p_t)^{N-i} \\ &= (1-r)Np_t + \frac{k_A}{N} (N^2 p_t^2 + Np_t(1-p_t)) + N - Np_t + k_B Np_t \\ &\quad - \frac{k_B}{N} (N^2 p_t^2 + Np_t(1-p_t)) \\ &= N + Np_t(k_B - r) + (k_A - k_B) (Np_t^2 + p_t(1-p_t)). \end{aligned} \quad (5.13)$$

E portanto, a Eq. (5.12) pode ser reescrita como

$$p_{t+1} = u + (1-2u) \frac{p_t(1-r) + k_A p_t^2 + \frac{1}{N} k_A p_t(1-p_t)}{1 + p_t(k_B - r) + (k_A - k_B) \left[\frac{1}{N} p_t(1-p_t) + p_t^2 \right]}. \quad (5.14)$$

5.2.1 Análise do modelo - Determinação de pontos fixos e cálculo de estabilidade

Vamos considerar primeiramente o caso da taxa de mutação nula ($u = 0$). Para este caso, podemos obter analiticamente as soluções de estado estacionário, $p_{t+1} = p_t = p^*$. Da Eq. (5.14), verificamos que os pontos fixos são soluções da equação $p^* - f(p^*) = 0$, onde

$$f(p^*) = \frac{p^*(1-r) + k_A p^{*2} + \frac{1}{N} k_A p^*(1-p^*)}{1 + p^*(k_B - r) + (k_A - k_B) \left[\frac{1}{N} p^*(1-p^*) + p^{*2} \right]}. \quad (5.15)$$

Explicitamente, determinamos três pontos fixos: $p^* = 0$, $p^* = 1$ e

$$p^* = \frac{r - k_A/N}{(k_A - k_B)(1 - 1/N)}. \quad (5.16)$$

Como mencionado anteriormente, um ponto fixo com algum sentido físico deve estar contido no simplex $[0, 1]$. No caso de equações de diferenças, a estabilidade de um ponto fixo é assegurada desde que [66, 67]

$$\left. \frac{dp_{t+1}}{dp_t} \right|_{p_t=p^*} < 1. \quad (5.17)$$

Verificamos que o ponto fixo $p^* = 0$ é estável para $k_A/r < N$, enquanto $p^* = 1$ é estável para $k_A/r > 1 + (1 - 1/N)k_B/r$. Interessantemente, para $k_B/r > N$ existe uma região onde ambos os pontos fixos são instáveis e o ponto fixo intermediário (Eq. (5.16)) é estável, que corresponde ao estado de coexistência entre altruístas e não-altruístas. Estes regimes distintos são ilustrados na Figura 5.1, onde mostramos as freqüências de estado estacionário p^* para dois valores distintos de freqüência inicial de altruístas. Podemos observar que no caso $u = 0$, a análise é consideravelmente simplificada pois apenas as razões $k_l/r, l = A, B$ têm importância para o cálculo da estabilidade dos pontos fixos. Identificamos quatro fases distintas no regime de estado estacionário:

- A fase completamente altruísta, (A) , associada com o ponto fixo $p^* = 1$;
- A fase completamente não-altruísta, (B) , associada com o ponto fixo $p^* = 0$;
- A fase de coexistência, (C) , associada ao ponto fixo (5.16);
- E a fase denominada $(A) - (B)$, onde ambos os pontos fixos $p^* = 1$ e $p^* = 0$ são estáveis.

Nesta última fase, os dois tipos de replicadores competem de forma que um domine completamente a população, embora o vencedor não seja determinado unicamente pelos seus valores seletivos, mas também pela sua freqüência inicial na população; sendo a bacia de atração dos dois pontos fixos estáveis delimitada pelo ponto fixo intermediário (5.16). Esses resultados são convenientemente resumidos em um diagrama de fases no plano $(k_A/r, k_B/r)$ como exibido na Figura 5.2(a). Podemos observar que as transições entre as fases (B) e (C) como também entre as fases (C) e (A) são contínuas, no sentido que p^* aumenta continuamente a medida em que cruzamos as linhas de transição. É importante observar que mesmo no caso de catálise não especificada ($k_A = k_B$), os replicadores altruístas podem dominar a população inteiramente, dada a condição $k_A > rN$.

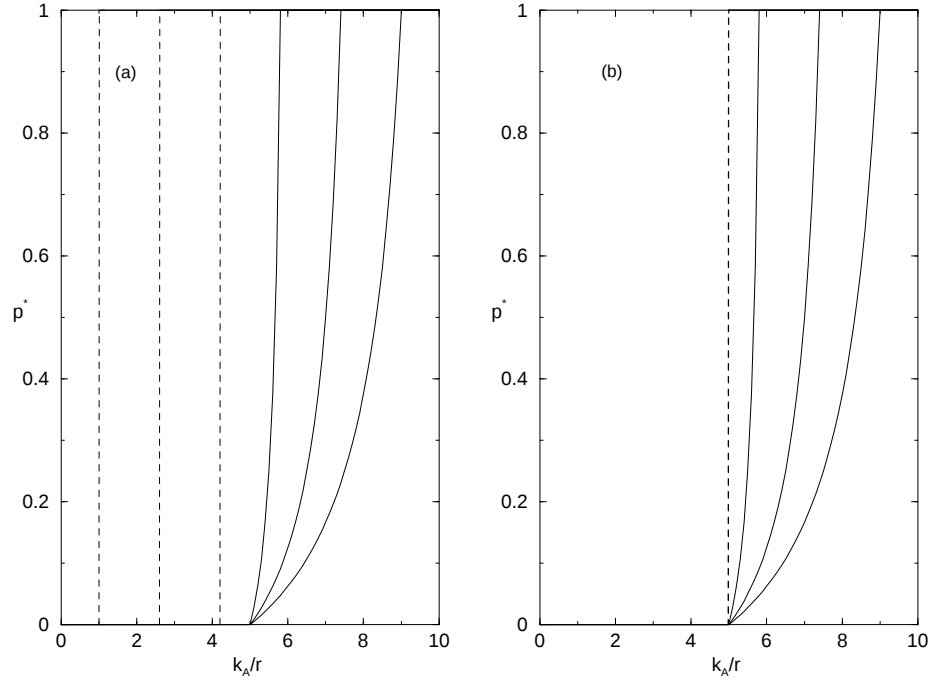


Figura 5.1: Frequência de estado estacionário de replicadores tipo A para $u = 0$, $N = 5$ e (da esquerda para direita) $k_B/r = 0, 2, 4, 6, 8$ e 10 . As frequências iniciais são (a) $p_0 = 0.999$ e (b) $p_0 = 0.001$. As três primeiras linhas na parte (a) colapsam em uma única curva na parte (b).

Analisaremos agora o caso de mutação não-nula. Para este caso, $p = 0$ e $p = 1$ não são mais soluções de estado estacionário da Eq. (5.14), e desta forma as fases não podem mais ser identificadas de maneira não-ambígua. Entretanto, o fenômeno de limiar observado na dependência da frequência de estado estacionário do replicador A em função da constante de catálise específica escalada k_A/r (ver Figura 5.3) indica que uma única extensão das definições de fases (A) , (B) e $(A) - (B)$ é possível, dado que k_B/r não seja maior que um determinado valor crítico. Como esperado, a fase (C) desaparece desde que sua definição característica, $0 < p^* < 1$, ocorre para todos os ajustes de parâmetro. A rica interação entre os pontos fixos estáveis é ilustrado nos diagramas de fases da Figura 5.2.

O aspecto mais notório desses diagramas é a existência de pontos críticos no qual as duas linhas de transição descontínuas interceptam-se, e como resultado, acima deste

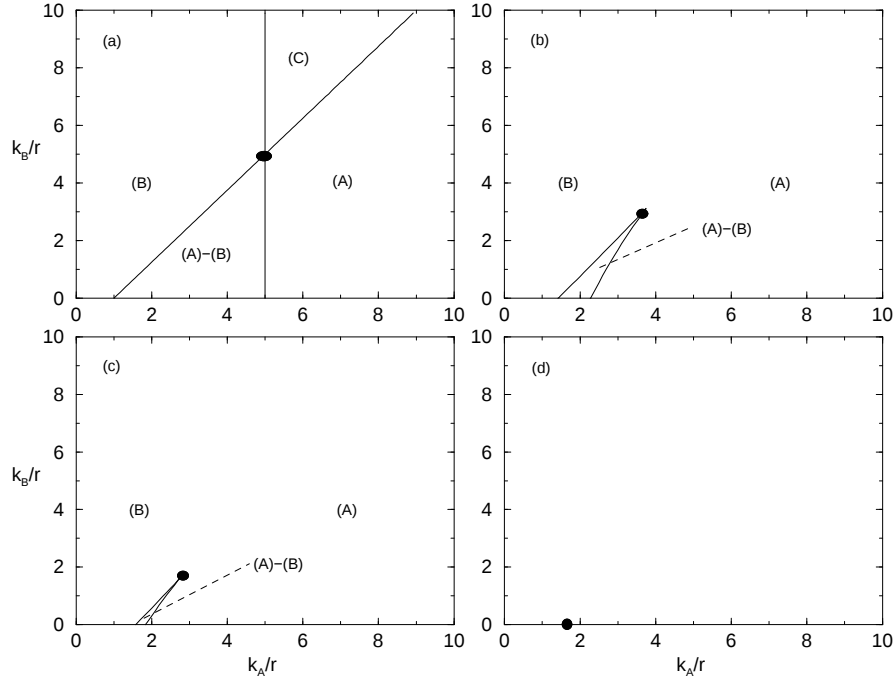


Figura 5.2: Diagramas de fases para $N = 5$ e $r = 0.1$ exibindo as regiões de estabilidade dos diferentes pontos fixos para (a) $u = 0$, (b) $u = 0.005$, (c) $u = 0.01$ e (d) $u = 0.0158$. O ponto de interseção toca o eixo coordenada no ponto $k_A = 5/3$.

ponto não é mais possível distinguir entre fases (A) e (B). Para valor fixo de u , r e N as coordenadas deste ponto (k_A^c, k_B^c) são determinadas por meio da condição que os três pontos fixos da equação de recorrência (5.14) colapse em um único ponto. De acordo com isto, exibimos na Figura 5.4 as coordenadas de ponto fixo como função da taxa de mutação u . Como esperado, para $u = 0$ encontramos que $k_B^c/r = N$ independentemente do valor de r . De particular interesse é a taxa de mutação em que k_B^c se anula, denotado por u_e , que sinaliza o desaparecimento de todos os traços dos dois regimes associados aos comportamentos altruísta e não-altruísta, levando ao diagrama de fases exibido na Figura 5.2 (d). Interessantemente, neste valor de taxa de mutação encontramos que $k_A^c = 2N/(N + 1)$ independentemente de r . A dependência de u_e com r , o custo em ser altruísta, para vários valores de tamanhos de demes N para a estabilização dos altruístas

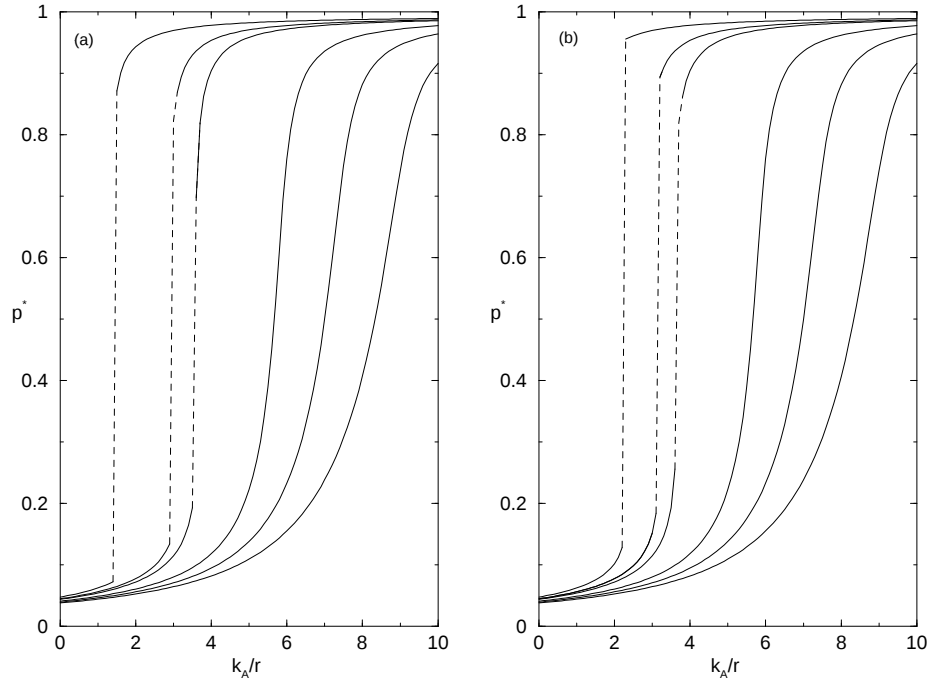


Figura 5.3: Frequência de estado estacionário de replicadores tipo A na população para $u = 0.005$, $N = 5$, $r = 0.1$, e (da esquerda para direita) $k_B/r = 0, 2.2, 9, 6, 8$ e 10 . As frequências iniciais são (a) $p_0 = 1$ e (b) $p_0 = 0$.

pode ser apreciada considerando o limite $N \rightarrow \infty$, que corresponde a uma população homogênea, quando trabalhamos com probabilidade de mutação nula, $u = 0$. De fato, neste caso o ponto fixo $p^* = 0$ é sempre estável, enquanto que $p^* = 1$ torna-se estável apenas para $k_A > r + k_B$, que é uma situação desinteressante do ponto de vista de evolução do altruísmo, pois o valor seletivo efetivo de um altruísta é maior que o valor seletivo de um elemento não-altruísta ($1 + k_B$) pertencendo a uma mesma *deme*.

5.2.2 Tamanho de demes variável

Nesta seção faremos uma breve análise para o caso em que as demes possuem tamanho variável. Vamos assumir que a distribuição para o tamanho das demes N obedece uma distribuição de Poisson, ou seja,

$$P(N) = \frac{e^{-\lambda} \lambda^N}{N!}, \quad (5.18)$$

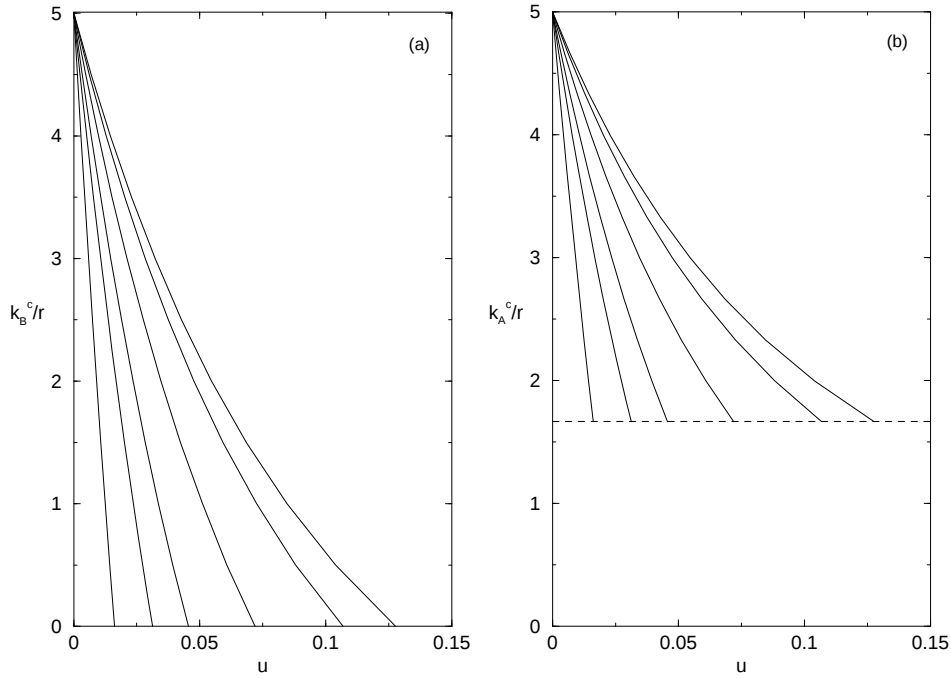


Figura 5.4: Coordenadas do ponto crítico (a) k_B^c/r e (b) k_A^c/r como funções da taxa de mutação u para $N = 5$ e (da esquerda para direita) $r = 0.1, 0.2, 0.3, 0.5, 0.8, e 1.0$. Nos pontos onde $k_B^c/r = 0$ encontramos $k_A^c/r = 5/3$.

com valor médio $\langle N \rangle = \lambda$.

Neste caso, o número médio de elementos altruístas $\mathcal{N}_A$ e não-altruístas $\mathcal{N}_B$ durante o estágio de crescimento ilimitado são iguais a

$$\mathcal{N}_A = \sum_{N=0}^{\infty} \sum_{i=0}^N [(1-u)iF_A(i) + u(N-i)F_B(i)] Y_t(i) \frac{e^{-\lambda} \lambda^N}{N!} \quad (5.19)$$

e

$$\mathcal{N}_B = \sum_{N=0}^{\infty} \sum_{i=0}^N [(1-u)(N-i)F_B(i) + uiF_A(i)] Y_t(i) \frac{e^{-\lambda} \lambda^N}{N!} \quad (5.20)$$

respectivamente, com $F_l(i), l = A, B$ e $Y_t(i)$ dados pela equações (5.10) e (5.5). Resolvendo os somatórios das equações (5.19) e (5.20), encontramos que a frequência global de altruístas na população na geração $t + 1$

$$p_{t+1} = \frac{\mathcal{N}_A}{\mathcal{N}_A + \mathcal{N}_B}, \quad (5.21)$$

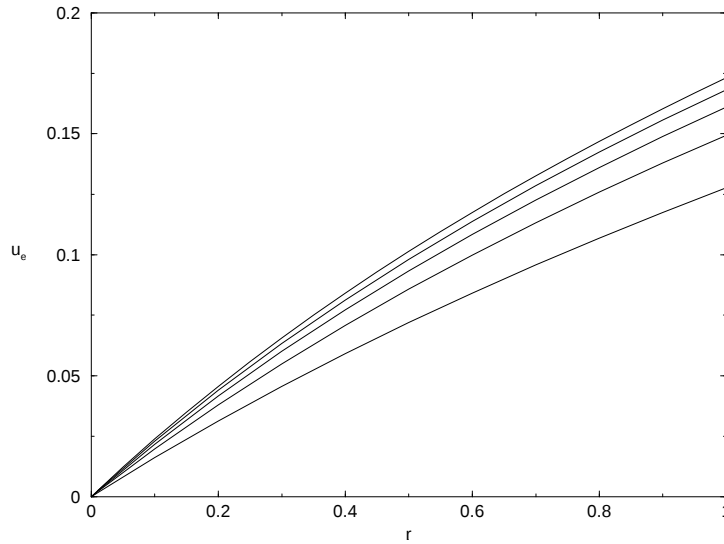


Figura 5.5: Taxa de mutação u_e além do qual as transições descontínuas desaparecem como função do custo altruísta r para (de baixo para cima) $N = 5, 10, 20, 50$ e ∞ .

resulta em

$$p_{t+1} = u + (1 - 2u) \frac{p_t(1 - r) + k_A p_t^2 + \frac{1}{\lambda} k_A p_t(1 - p_t)}{1 + p_t(k_B - r) + (k_A - k_B) \left[\frac{1}{\lambda} p_t(1 - p_t) + p_t^2 \right]}, \quad (5.22)$$

que é idêntico ao resultado obtido anteriormente para o caso de demes de tamanho fixo, com λ substituindo N . Portanto, todos os resultados e análise de pontos fixos obtidos anteriormente são igualmente válidos para o caso de demes de tamanhos variável com distribuição de Poisson.

5.3 Sinergia

Um problema intrigante em evolução é a existência de estruturas complexas que são de grande valor para os organismos quando completamente formados [15]. Possivelmente, a produção de enzimas tenha-se tornado uma realidade devido à ação combinada de várias moléculas, cada uma sendo responsável pela síntese de diferentes partes do catalizador. Esta divisão de tarefas entre os altruístas, denominada de sinergia, pode resultar em interações adaptativas não-aditivas altamente não-lineares. De forma a modelarmos esta situação, assumimos que uma certa vantagem seletiva é conferida a *deme* quando o número

de altruístas alcança um valor mínimo i_m . Assumimos que as demes compostas por $i \geq i_m$ elementos altruístas, onde $i_m = 0, 1, \dots, N$, terão seus valores seletivos incrementados por um fator $1/(1-c)$, com $c \in [0, 1]$. Desta forma, os valores seletivos dos elementos altruístas e não altruístas são dados por:

Tabela 1: Valores seletivos dos elementos na população

	$i < i_m$	$i \geq i_m$
altruísta	$1 - r$	$(1 - r)/(1 - c)$
não-altruísta	1	$1/(1 - c)$

onde mais uma vez r é o custo associado ao elemento altruísta, e c é denominado de taxa de seleção interdêmica.

Das Eqs. (5.1) e (5.4), podemos ver que a equação de recorrência para a fração de elementos altruístas na população é dada por

$$\begin{aligned}
 p_{t+1} &= \frac{1}{w_{tot}} \left\{ \sum_{i=0}^{i_m-1} [(1-r)i(1-u) + (N-i)u] Y_t(i) \right. \\
 &\quad \left. + \sum_{i=i_m}^N \left[\frac{(1-r)}{(1-c)} i(1-u) + \frac{1}{(1-c)} (N-i)u \right] Y_t(i) \right\} \\
 &= \frac{1}{w_{tot}} \left\{ \frac{u}{1-c} \left[(1-c)N(1-rp_t) + Nc \sum_{i=i_m}^N Y_t(i) \left(1 - \frac{1}{N} r i \right) \right] \right. \\
 &\quad \left. + \frac{1-2u}{1-c} (1-r) \left[Np_t(1-c) + c \sum_{i=i_m}^N i Y_t(i) \right] \right\} \quad (5.23)
 \end{aligned}$$

e a taxa média de reprodução da população é igual a

$$\begin{aligned}
 w_{tot} &= \sum_{i=0}^{i_m-1} [(1-r)i + (N-i)] \binom{N}{i} p_t^i (1-p_t)^{N-i} + \\
 &\quad \sum_{i=i_m}^N \left[\frac{(1-r)}{(1-c)} i + \frac{(N-i)}{(1-c)} \right] \binom{N}{i} p_t^i (1-p_t)^{N-i} \\
 &= N(1-rp_t) + \frac{Nc}{1-c} \left[\sum_{i=i_m}^N Y_t(i) \left(1 - \frac{1}{N} r i \right) \right]. \quad (5.24)
 \end{aligned}$$

E então, a Equação (5.23) pode ser reescrita como

$$p_{t+1} = u + (1 - 2u) \frac{(1 - r) \left[p_t(1 - c) + \frac{1}{N} c \sum_{i=i_m}^N i Y_t(i) \right]}{(1 - c)(1 - r p_t) + c \sum_{i=i_m}^N Y_t(i) (1 - \frac{1}{N} r i)}, \quad (5.25)$$

com

$$Y_t(i) = \binom{N}{i} p_t^i (1 - p_t)^{N-i}. \quad (5.26)$$

Procederemos agora com a análise de estado estacionário da Equação (5.25).

Para ressaltar, os valores seletivos atribuídos aos elementos da população e definidos na Tabela 1, foram utilizadas por Donato [18] em um modelo alternativo para seleção do comportamento altruísta, que semelhantemente ao modelo de Wilson de *demes* estruturadas, existe um estágio do ciclo de vida dos indivíduos onde estes interagem com todos os outros indivíduos da população. Entretanto, o modelo proposto por Donato et al. [18, 68] possui dois ingredientes distintos:

- O tamanho das *demes* não é fixo, mas há um tamanho máximo a ser alcançado, e neste ponto a *deme* é dividida em duas novas *demes* menores.
- Há sobreposição de gerações.

Essas regras foram motivadas por meio de analogias com grupos de animais, que vivem em grupos não tão grandes e cujos filhos permanecem no mesmo grupo de seus pais. Um resultado interessante do modelo é a possibilidade de coexistência estável entre altruístas e não-altruístas dentro de um mesmo grupo; contradizendo resultados para o modelo de seleção de grupo de ilhas, onde na ausência de mutações apenas os grupos completamente altruístas ($i = N$) ou completamente não-altruístas ($i = 0$) estão presentes [60–62].

5.3.1 Análise do modelo - Determinação de pontos fixos e cálculo de estabilidade

Consideraremos primeiramente o caso mais simples, onde $u = 0$ (sem mutação). Para este caso a Equação (5.25) iguala a

$$p_{t+1} = \frac{(1-r) \left[p_t(1-c) + \frac{1}{N}c \sum_{i=i_m}^N iY_t(i) \right]}{(1-c)(1-rp_t) + c \sum_{i=i_m}^N Y_t(i) \left(1 - \frac{1}{N}ri\right)}. \quad (5.27)$$

A equação acima tem sempre como pontos fixos os valores $p = 0$ e $p = 1$. Dependendo dos parâmetros de controle do modelo i_m , c e r , pode haver um ou dois pontos fixos adicionais. Verificamos que a solução $p^* = 0$ é sempre estável para $i_m > 1$, enquanto que para $i_m = 1$ ela torna-se instável na região $c > r$. Para estes casos em particular, podemos facilmente realizar o cálculo da estabilidade. Para $i_m > 1$, encontramos que

$$\left. \frac{dp_{t+1}}{dp_t} \right|_{p^*=0} = 1 - r, \quad (5.28)$$

e desde que $r \in [0, 1]$, o ponto fixo $p^* = 0$ é sempre estável. Já Para $i_m = 1$, a Equação (5.27) iguala a

$$p_{t+1} = \frac{(1-r)Np_t}{(1-rp_t) - c(1-p_t)^N}, \quad (5.29)$$

e neste caso temos

$$\left. \frac{dp_{t+1}}{dp_t} \right|_{p^*=0} = \frac{1-r}{1-c}. \quad (5.30)$$

Logo a estabilidade da solução é verificada para $c < r$. De fato, para valor fixo de r , um ponto fixo estável aparece para $c = r$, aumentando continuamente a medida em que incrementamos c . Este comportamento sinaliza a ocorrência de uma transição contínua de um regime caracterizado por demes completamente não-altruístas ($p^* = 0$) para um regime onde demes inhomogêneas, formadas por indivíduos altruístas e não-altruístas, coexistem ($0 < p^* < 1$) (ver Figura 5.6).

Já para $p^* = 1$ e $i_m < N$ encontramos que

$$\left. \frac{dp_{t+1}}{dp_t} \right|_{p^*=1} = 1 + \frac{r}{1-r} > 1, \quad (5.31)$$

indicando que a solução é sempre instável.

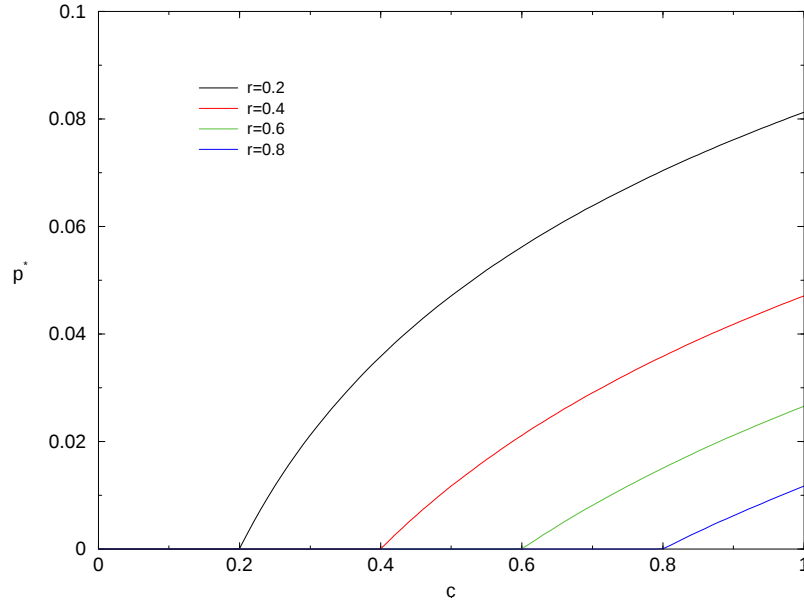


Figura 5.6: Estado estacionário p^* da equação (5.27) em função de c , para alguns valores de r , e valores fixos de $N = 20$ e $i_m = 1$.

O caso em que $i_m = N$ pode ser facilmente analisado: a equação de recorrência (5.25) é reescrita como

$$p_{t+1} = \frac{p_t(1-r) - c(1-r)(p_t - p_t^N)}{(1-rp_t) + c[r(p_t - p_t^N) - (1-p_t^N)]}, \quad (5.32)$$

onde encontramos para $p^* = 1$ que

$$\left. \frac{dp_{t+1}}{dp_t} \right|_{p^*=1} = \frac{1-c}{1-r}, \quad (5.33)$$

e portanto a solução é estável na região $c > r$. Neste caso, verificamos que ambos os pontos fixos $p^* = 0$ e $p^* = 1$ são estáveis, porém apenas um dos dois tipos de indivíduos povoará toda a população. Nos casos intermediários $1 < i_m < N$, pontos fixos estáveis $0 < p^* < 1$ surgem de forma descontínua.

Na Figura 5.7 apresentamos as linhas de transição separando a região no plano (c, r) onde os indivíduos altruístas persistem na população (região abaixo da curva) da região onde o único ponto fixo estável é o ponto não-altruísta $p^* = 0$. Como essas curvas

satisfazem a condição $c \geq r$ para qualquer valor de i_m , isto parece indicar que os altruístas sobreviventes são aqueles que pertencem à *demes* com $i \geq i_m$ elementos altruístas, pois estes possuem um maior valor seletivo que os indivíduos não-altruístas que vivem em *demes* com $i < i_m$. Podemos verificar também que o tamanho da região de existência dos altruístas decresce com o aumento de i_m , alcança um valor mínimo para $i_m = N/2$, e então aumenta novamente em direção ao seu tamanho inicial a medida em que i_m se aproxima de N (as linhas de transição para $i_m = 1$ e $i_m = N$ coincidem). Entretanto, deve ser observado que a situação mais favorável aos altruístas ocorre para o caso de nenhuma sinergia $i_m = 1$, desde que então o ponto fixo $p^* = 0$, associado com o regime não-altruísta, torna-se instável. Além do mais, a bacia de atração do ponto fixo intermediário decresce com o aumento de i_m , e desta forma, ao menos que exista em princípio uma grande número de indivíduos altruístas na população, os não-altruístas dominarão toda a população. Por exemplo, para $i_m = N$ a bacia de atração de $p^* = 1$ é muito pequena próxima a linha de transição $c = r$. Este resultado frustrante simplesmente reflete a dificuldade de evolução de um sistema sinérgico na natureza.

Quando levamos em consideração o efeito de mutação ($u > 0$) no sistema, isto nos leva a uma interação muito mais rica entre os diferentes regimes de estado estacionário da equação de recorrência (5.25), como ilustrado pelo diagramas de fases exibidos na Figura 5.8. Na ausência de mutação, a fase rotulada (B) é associada com o regime não-altruísta, caracterizada pelo ponto fixo $p^* = 0$; enquanto que a fase (C) é associada com a região de coexistência caracterizada pelo ponto fixo intermediário $0 < p^* < 1$. Embora para taxa de mutação não-nula, $p = 0$ não seja mais ponto fixo, é ainda possível distinguir entre os pontos fixos correspondendo aos regimes não-altruísta e o de coexistência, devido à ocorrência de fenômenos de limiar semelhantes àqueles vistos na Figura 5.3. O principal efeito da mutação é produzir, às custas da fase (C), uma região limitada, denominada (B)-(C), onde ambas as soluções são estáveis. Essa região é delimitada por duas linhas

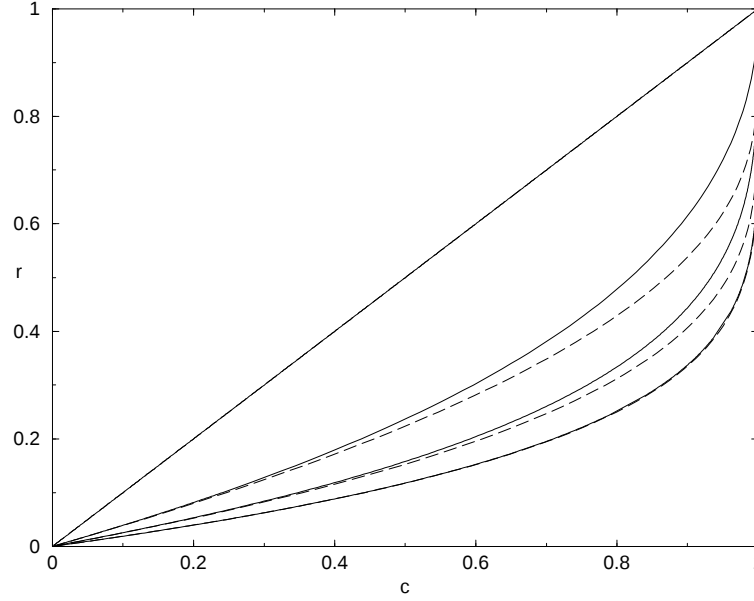


Figura 5.7: Linhas de transição para $N = 20$, $u = 0$ e, de cima para baixo, $i_m = 1, 2, 4, 10$ (linhas sólidas), e $i_m = 19, 17$ (linhas tracejadas). As curvas para $i_m = 20$ e $i_m = 11$ coincidem com aqueles obtidos para $i_m = 1$ e $i_m = 10$, respectivamente.

de transição descontínuas que se cruzam e que terminam em dois pontos críticos. À medida em que incrementamos a taxa de mutação u , o tamanho da região é reduzido e desaparece completamente no ponto crítico u_e , em que os dois pontos colapsam. Portanto, para $u \geq u_e$ não é mais possível distinguir entre fases (B) e (C). A dependência de u_e em i_m é exibida na Figura 5.9. Como esperado, $u_e = 0$ para $i_m = 1$, independentemente do tamanho da *deme* N , pois a transição entre as duas fases já é contínua para $u = 0$.

5.3.2 Tamanho de demes variável

Nesta seção, faremos uma análise simplificada para o caso em que o tamanho das *demes* que compõem a população são de tamanho variável. Assumimos que a distribuição de tamanhos das *demes* obedece uma distribuição de Poisson, ou seja,

$$P(N) = \frac{e^{-\lambda} \lambda^N}{N!}. \quad (5.34)$$

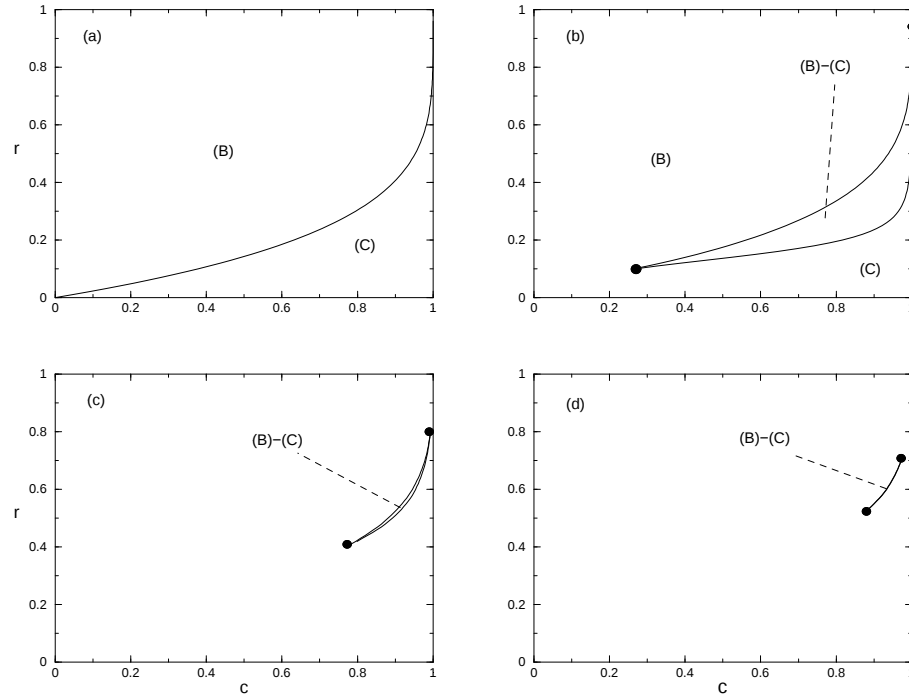


Figura 5.8: Diagrama de fases para $N = 20$ e $i_m = 5$ exibindo as regiões de estabilidade dos diferentes pontos fixos para (a) $u = 0$, (b) $u = 0.0085$, (c) $u = 0.0282$ e (d) $u = 0.0330$.

Como mencionado anteriormente, em sua formulação original, Donato et al. [18] consideraram *demes* de tamanho variável, embora naquele trabalho exista um valor máximo permitido, a partir do qual a *deme* é dividida. Considerando esta nova formulação, o número médio de indivíduos altruístas $\mathcal{N}_A$ e de não-altruístas $\mathcal{N}_B$ são dados por

$$\mathcal{N}_A = \sum_{N=0}^{\infty} \frac{e^{-\lambda} \lambda^N}{N!} \left[\sum_{i=0}^{x_c N-1} \left[(1-u)i(1-r) + u(N-i) \right] Y_t(i) + \sum_{i=x_c N}^N \left[(1-u)i \frac{(1-r)}{(1-c)} + u(N-i) \frac{1}{(1-c)} \right] Y_t(i) \right] \quad (5.35)$$

e

$$\mathcal{N}_B = \sum_{N=0}^{\infty} \frac{e^{-\lambda} \lambda^N}{N!} \left[\sum_{i=0}^{x_c N-1} \left[(1-u)(N-i) + ui(1-r) \right] Y_t(i) + \sum_{i=x_c N}^N \left[(1-u)(N-i) \frac{1}{(1-c)} + ui \frac{(1-r)}{(1-c)} \right] Y_t(i) \right] \quad (5.36)$$

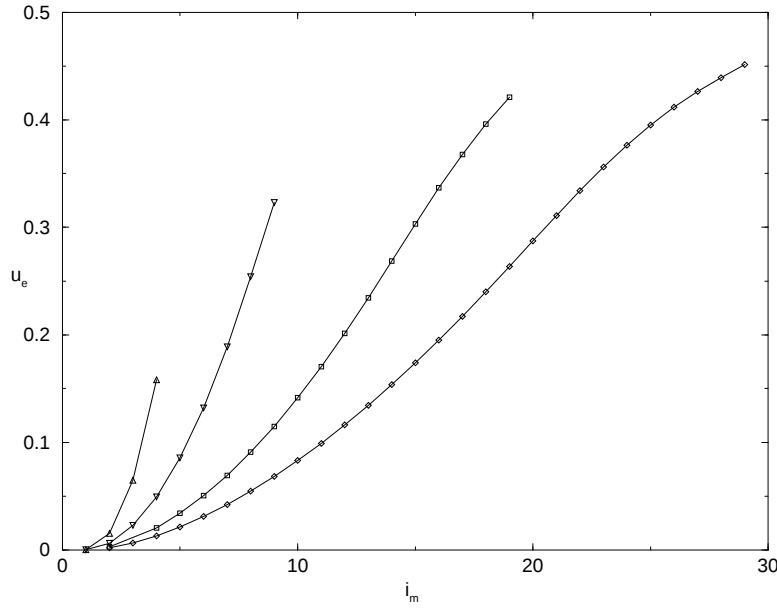


Figura 5.9: Taxa de mutação u_e , além do qual as transições descontínuas desaparecem, como função de i_m para $N = 5$ ($\triangle$), $N = 10$ (∇), $N = 20$ ($\square$) e $N = 30$ ($\diamond$).

respectivamente, onde $x_c \in [0, 1]$ descreve o valor crítico de concentração de altruístas, a partir do qual os valores seletivos dos indivíduos que compõem a *deme* têm seus valores seletivos amplificados por um fator $1/(1 - c)$, e $Y_t(i)$ é descrita pela Eq. (5.5). Uma investigação analítica e mesmo numérica de regime estacionário $p_{t+1} = p_t = p^*$ da equação de recorrência descrevendo a concentração de altruístas

$$p_{t+1} = \frac{\mathcal{N}_A}{\mathcal{N}_A + \mathcal{N}_B} \quad (5.37)$$

é bastante difícil. Vamos nos restringir aqui a análise de casos especiais. Vamos nos concentrar unicamente no caso de taxa de mutação nula $u = 0$ e na análise de estabilidade das soluções $p^* = 0$ e $p^* = 1$.

Para o caso de mutação nula, a Eq. (5.35) se reduz a

$$\mathcal{N}_A = \sum_{N=0}^{\infty} \frac{e^{-\lambda} \lambda^N}{N!} \left[(1 - r)(Np_t - \sum_{i=x_c N}^N iY_t(i)) + \frac{(1 - r)}{(1 - c)} \sum_{i=x_c N}^N iY_t(i) \right] \quad (5.38)$$

e

$$\begin{aligned} w_{tot} &= \mathcal{N}_A + \mathcal{N}_B \\ &= \sum_{N=0}^{\infty} \frac{e^{-\lambda} \lambda^N}{N!} \left[N(1 - rp_t) + \frac{c}{1-c} \left[N \sum_{i=x_c N}^N Y_t(i) - r \sum_{i=x_c N}^N iY_t(i) \right] \right]. \end{aligned} \quad (5.39)$$

O caso $x_c = 1$

Substituindo $x_c = 1$ nas Eqs. (5.38) e (5.39), encontramos que a equação de recorrência para a frequência de altruísta na população é

$$p_{t+1} = \frac{(1-r)p_t + \frac{c}{1-c}(1-r)p_t e^{-\lambda(1-p_t)}}{(1-rp_t) + \frac{c}{1-c}(1-r)p_t e^{-\lambda(1-p_t)}}. \quad (5.40)$$

Explicitamente, encontramos três pontos fixos: $p^* = 0$, $p^* = 1$ e

$$p^* = 1 + \frac{1}{\lambda} \ln \left[\frac{r(1-c)}{c(1-r)} \right]. \quad (5.41)$$

Como discutido anteriormente, um ponto fixo com algum sentido físico, deve estar contido no simplex $[0, 1]$ e satisfazer a condição de estabilidade

$$\left. \frac{dp_{t+1}}{dp_t} \right|_{p_t=p^*} < 1. \quad (5.42)$$

Diferentemente do resultado obtido para *demes* de tamanho fixo, onde o ponto fixo $p^* = 0$ é sempre estável quando $i_m = N$, encontramos que existe uma pequena região em que esta solução torna-se instável. A condição de estabilidade para $p^* = 0$ é dada por

$$c < \frac{r}{e^{-\lambda} - r(e^{-\lambda} - 1)}. \quad (5.43)$$

No limite $\lambda \rightarrow \infty$, a solução é estável em todo o espaço de fases. Já para $p^* = 1$, a condição de estabilidade $c > r$ é igual à obtida anteriormente quando consideramos *demes* de tamanho fixo. A solução intermediária (5.41) é sempre instável na região de interesse.

O caso x_c geral

Para este caso particular vamos nos restringir unicamente a análise dos pontos fixos $p^* = 0$ e $p^* = 1$. Para a solução completamente não-altruísta $p^* = 0$, a região de estabilidade é dada por

$$c < \frac{r}{e^{-\lambda} - r(e^{-\lambda} - 1)} , \quad (5.44)$$

exatamente idêntica à obtida quando $x_c = 1$. Já a solução altruísta $p^* = 1$ é sempre instável. Mesmo resultado verificado para as *demes* de tamanho fixo.

Capítulo 6

Evolução em Relevos Rugosos

Neste capítulo apresentamos alguns resultados e estudos apresentados na literatura sobre evolução e adaptação em relevos rugosos. Este tipo de relevo é mais adequado para descrever relevos reais encontrados na natureza. Veremos que alguns aspectos inesperados surgem quando lidamos com relevos com propriedades distintas das que exibimos até então.

Relevos de energia com muitos mínimos locais são um tópico de muito estudo na mecânica estatística de vidros de spin, e também são bastante estudados atualmente no contexto de enovelamento de RNA e proteínas (ver Fig. 6.2). Relevos similares de *adaptação*, com muitos máximos locais (ao invés de mínimos), têm atraído a atenção em biologia evolucionária e em ciência computacional. Biólogos teóricos utilizam esses relevos em modelos de evolução. Cientistas em computação também lidam com esses relevos, desde que estes surgem em problemas de otimização combinatorial, e no treinamento de redes neurais.

Na próxima seção, lidaremos mais uma vez com o modelo de quase-espécies, porém utilizando um outro *relevo de adaptação*. Nosso objetivo é tentar mostrar como algumas propriedades de relevos reais podem ser incorporadas a este modelo bastante simples, e ver suas consequências. Na segunda seção descrevemos um modelo que gera uma família de relevos rugosos, denominado modelo NK.

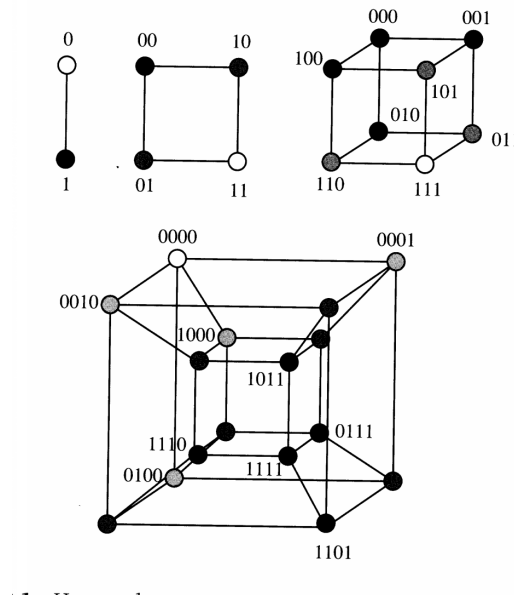


Figura 6.1: Hipercubos representando o espaço de seqüências para seqüências binárias de comprimento de 1 a 4.

6.1 Redes neutras e competição entre diferentes quase-espécies

Uma característica importante verificada em relevos reais como os de RNA e proteínas é a existência de um alto grau de degenerescência no mapeamento genótipo-fenótipo (ver Fig. 6.3). Ou seja, mais de um genótipo é mapeado em uma mesma estrutura, e portanto desempenha uma mesma função. A noção de evolução neutra foi criada por Motoo Kimura [69, 70] de forma a levar em consideração o fato bem estabelecido de que quase todas as mutações registradas a nível de seqüências de DNA na natureza são seletivamente neutras. A maioria das mutações ocorrem em freqüências que são aparentemente independentes das alterações morfológicas observadas nas espécies. Em essência, a teoria neutra reflete as relações entre genótipos e fenótipos e confirma que uma fração substancial das mutações pontuais não alteram a adaptação. Ela assume implicitamente que o papel primário da seleção é eliminar as variantes deletéreas.

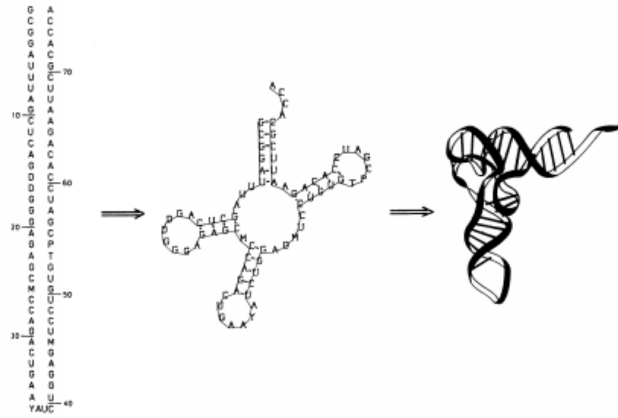


Figura 6.2: Enovelamento de uma sequência de RNA em sua estrutura espacial.

A motivação para o estudo de relevos adaptativos com neutralidade tem crescido nos últimos anos, principalmente com os avanços adquiridos na área de Biologia Molecular. Foi verificado que coleções de genótipos mutualmente neutros, que são conectados por meio de um único passo de mutação, formam redes prolongadas que permeiam grandes regiões do espaço de genótipos. Intuitivamente, uma grande degenerescência no mapeamento genótipo-fenótipo, quando combinadas com a alta conectividade do espaço de genótipos, prontamente nos leva a tais redes neutras prolongadas. Essa intuição é sustentada por resultados teóricos recentes [71, 72]. Na evolução *in vitro* de ribosimas, mutações responsáveis por um incremento no valor seletivo são apenas uma pequena minoria do número total de mutações aceitas [73]. Este fato indica que, mesmo em evolução adaptativa, a maioria das mutações pontuais são neutras.

O fato que apenas uma minoria dos loci é conservada em seqüências que evoluíram de um único ancestral indica também uma alta degenerescência no mapeamento genótipo-fenótipo ribosimal [74]. Neutralidade também é constatada em experimentos onde seqüências de RNA evoluíram para uma dada estrutura iniciando de um número de diferentes genótipos iniciais [75]. De forma mais geral, neutralidade em mapeamentos genótipo-fenótipo em RNA e proteínas é indicada pela observação que suas estruturas são

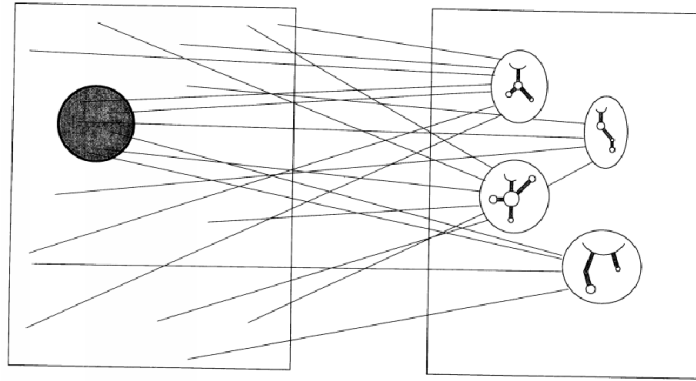


Figura 6.3: Um esboço do mapeamento de seqüências de RNA em suas estruturas secundárias.

muito mais conservadas durante a evolução que suas seqüências primárias.

Um outro tipo de abordagem para a dinâmica evolucionária é sua modelagem por meio de uma plataforma artificial de evolução. Muitas questões fundamentais em Biologia são difíceis de serem enfocadas devido a alta dimensionalidade dos genomas e como também às dificuldades práticas de manipulação de numerosos genótipos e análise de suas propriedades fenotípicas. Algum progresso tem sido realizado no estudo de microorganismos, mas estes problemas são ainda desanimadores [76, 77]. Uma abordagem alternativa envolve o estudo de vida artificial, em particular *códigos de computadores* ou *organismos digitais* que compartilham com os organismos reais as propriedades de auto-replicação e mutação, como também genomas com alta dimensionalidade e portanto trajetórias evolucionárias indeterminadas. Os resultados dos experimentos aqui utilizados foram realizados utilizando a versão 1.4 do programa *Avida*, uma plataforma flexível para pesquisa em vida artificial [78, 79]. De forma breve, *organismos digitais* são códigos de programas auto-replicantes que competem por tempo da unidade de processamento central (CPU), que no caso é o *combustível* necessário para replicação. Os programas mutam de forma aleatória e evoluem em um meio computacional bem definido. Cada

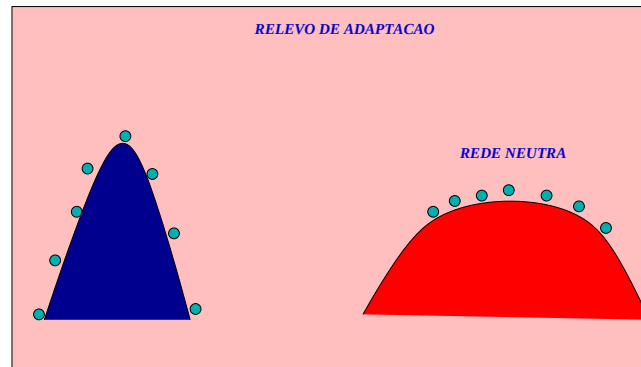


Figura 6.4: Um esboço do relevô de adaptação em organismos digitais.

organismo possui um genoma cujo comprimento é medido como o número de instruções seqüenciais em seu código. A seqüência de instruções pode ser alterada por mutação, incluindo eventos de inserção e deleção como também mutações que alteram uma instrução por outra. Existem 28 instruções distintas, que podem ser imaginadas como o análogo dos 20 diferentes amino-ácidos presentes nas proteínas. Os resultados e modelagem aqui realizados foram desenvolvidos em colaboração com o Dr. Claus Wilke e Prof. Dr. Chris Adami do *Digital Life Laboratory no California Institute of Technology (CALTECH)*.

Nesta seção estamos interessados em investigar a competição entre duas quase-espécies em duas redes neutras distintas. Consideremos que essas duas redes possuem topologias diferentes como também valores seletivos distintos. Essa análise é motivada principalmente para tentar compreender a dinâmica evolucionária nos relevos reais, desde que devido à rugosidade extremamente elevada como também devido à existência de redes neutras nesses relevos, esse cenário é bastante comum (ver Figura 6.4). De forma a modelarmos essa situação trabalharemos com o modelo de quase-espécies.

6.1.1 Relevô de dois picos com neutralidade

Como modelo teórico, consideramos o modelo de quase-espécies descrito no capítulo 2. Como vimos anteriormente, a dinâmica do modelo é descrita pelo conjunto de equações

diferenciais (2.6)

$$\frac{dY_P}{dt} = \sum_{R=0}^L A'_R M_{PR} Y_R - \Phi(t) Y_P. \quad (6.1)$$

onde Y_P denota a concentração da classe P de seqüências e M_{PR} é a matriz de mutação da classe R para classe P . De forma a modelarmos a situação descrita na Figura 6.4, utilizamos aqui o relevo de dois picos com neutralidade. Na Figura 6.5 exibimos o relevo aqui utilizado para obtenção de nossas predições teóricas. Para obtenção dos resultados aqui exibidos utilizamos seqüências de tamanho $L = 20$. O relevo de adaptação apresenta dois picos: O primeiro pico consiste de todas as seqüências com distância de Hamming menor que 3 em relação à seqüência 000...000, e tem valor seletivo igual a 7; e o segundo consistindo de todas as seqüências com distância de Hamming menor que 5 em relação à seqüência 111...111, e possuem valor seletivo 4. Diferentemente do relevo utilizado no capítulo 2, onde todas as seqüências exceto a mestra têm valor seletivo 1 e portanto são capazes de se auto-reproduzirem, aqui consideramos que uma fração substancial dos tipos de seqüências não possuem essa capacidade, ou seja, têm taxa de reprodução nula. Essa suposição é plenamente justificável desde que em relevos reais uma fração das seqüências não desenvolvem qualquer função, são consideradas seqüências mortas.

Portanto, consideramos que o relevo de replicação exibe um certo grau de neutralidade. E conforme o nosso exemplo, o grau de neutralidade é maior para o segundo pico.

Na Figura 6.6 exibimos os valores de estado estacionário para o valor seletivo médio e máximo da população, a concentração das seqüências nos dois picos do relevo e a concentração de seqüências mortas, todos em função do parâmetro $(1 - q)$, probabilidade de mutação por dígito da seqüência. Como podemos observar, existe uma transição abrupta no comportamento do sistema. O valor seletivo médio da população apresenta um decaimento contínuo e praticamente não nos dá qualquer informação relevante sobre

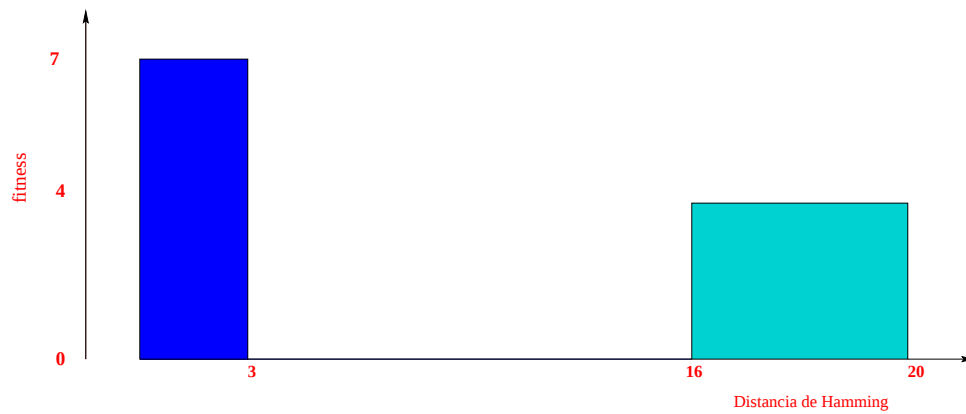


Figura 6.5: Relevo de adaptação de dois picos. Aqui utilizamos seqüências de tamanho $L = 20$. O primeiro pico tem valor seletivo 7 com largura igual 3, ou seja consiste de todas as seqüências com distância de Hamming menor que 3 em relação à seqüência 000...000. Já as seqüências no segundo pico têm valor seletivo 4 e consiste de todas as seqüências com distância de Hamming menor que 5 em relação à seqüência 111...111.

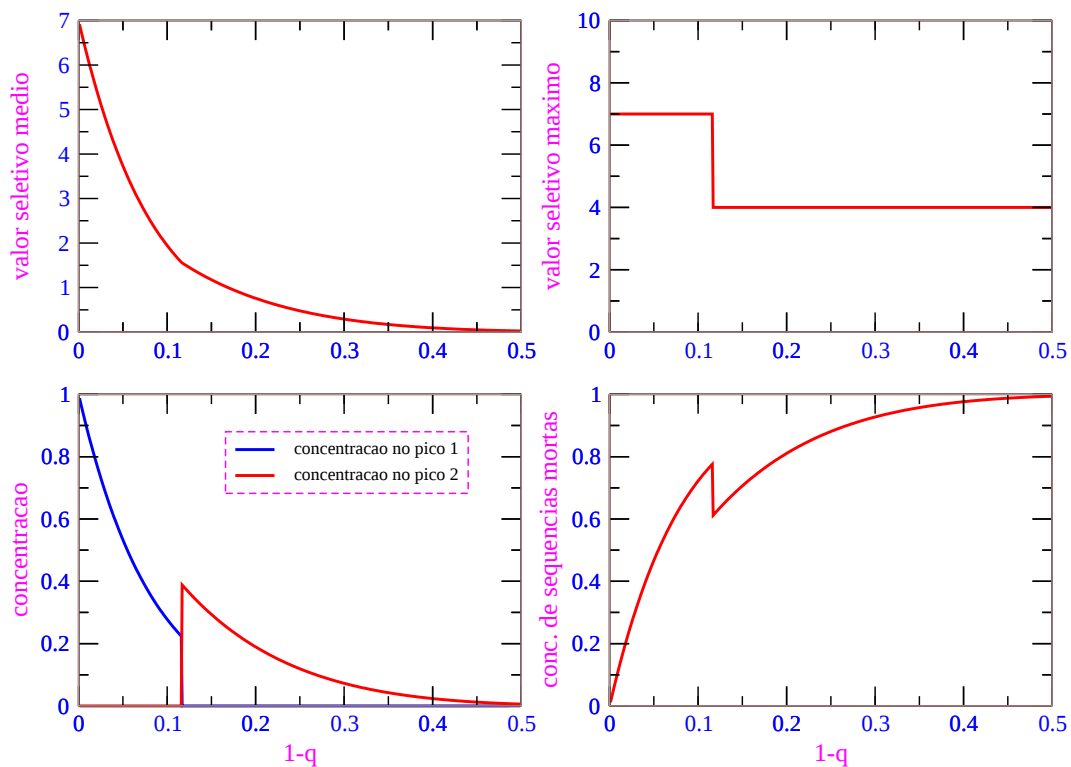


Figura 6.6: (a) valor seletivo médio, (b) valor seletivo máximo, (c) concentração de seqüências nos picos 1 e 2 do relevo, e por fim (d) concentração de seqüências mortas; todos como função do parâmetro $(1 - q)$.

o que acontece com o sistema. Exceto por um ponto de descontinuidade em sua derivada no ponto de transiç o. Entretanto, podemos ver mais claramente das Figuras 6.6 (b). (c) e (d) uma mudan a de comportamento em torno do ponto $1 - q \approx 0.1$. Neste ponto, a quase-esp cie situada no segundo pico passa a dominar, excluindo da popula o a quase-esp cie situada no primeiro pico do relevô. Abaixo desse ponto, a popula o   inteiramente povoada por seq ncias situadas no primeiro pico do relevô e por seq ncias mortas.

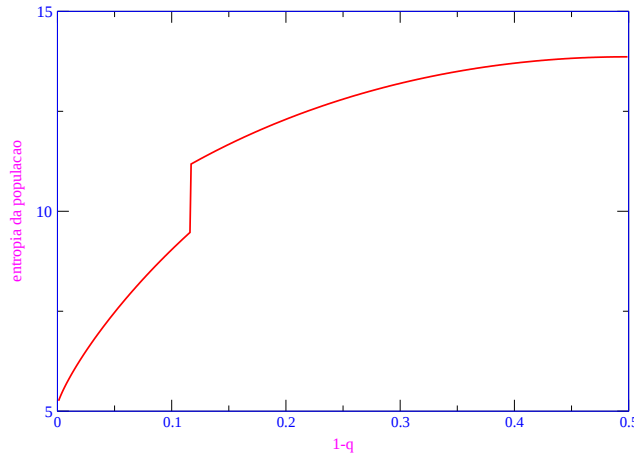


Figura 6.7: Entropia da popula o como fun o do par metro $(1 - q)$.

Essa mudan a de comportamento no sistema pode ser visualizada tamb m atrav s da an lise da entropia de Shannon [80, 81], definida por

$$S(\mathbf{x}) = - \sum_i x_i \ln x_i, \quad (6.2)$$

onde o somat rio se d  sobre todos os tipos de seq ncias. Na Figura 6.7 podemos claramente visualizar um aumento abrupto da entropia. Esse aumento se deve ao dom nio da popula o pela quase-esp cie situada no segundo pico. Desde que o grau de neutralidade nessa segunda rede   maior que na primeira, existe um aumento do n mero de graus de liberdade no sistema.

Nas Figuras 6.8 e 6.9 exibimos os resultados de nossa simula o para os organismos digitais. Para a obten o desses resultados consideramos popula es de tamanho

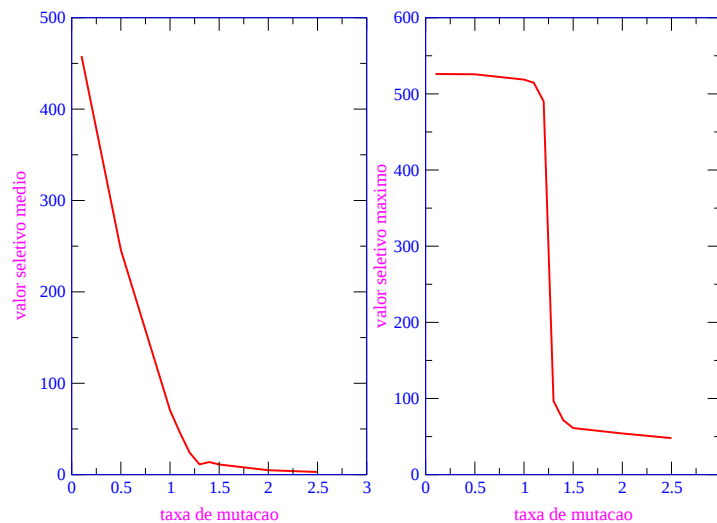


Figura 6.8: Resultados da simulação para organismos digitais. Temos na parte (a) Valor seletivo médio e (b) Valor seletivo máximo da população com função da taxa de mutação.

fixo e igual a 3600, e o número de amostras realizadas igual a 10. Como podemos observar, registramos o mesmo tipo de comportamento para o modelo teórico descrito. Na Figura 6.8 podemos observar de forma mais pronunciada o ponto de descontinuidade na derivada do valor seletivo médio. Isto deve-se provavelmente a efeitos de finitude da população em nossa simulação. Uma nítida transição é visualizada através do gráfico do valor máximo seletivo presente na população como função da taxa de mutação. Através da entropia também podemos visualizar claramente essa transição, embora ela seja muito mais evidente quando mostramos a exponencial da entropia como função da taxa de mutação.

Podemos concluir com esses resultados que, sob a suposição de que duas redes neutras têm topologias distintas e diferentes níveis seletivos, é a taxa de mutação que determina que quase-espécie eventualmente será levada à extinção. Para pequenos valores de taxa de mutação, encontramos que a quase-espécie residindo na rede neutra com menor taxa de replicação será eliminada. Já para grandes valores de taxa de mutação, as seqüências mais eficientes em replicar-se são excluídas pelas menos eficientes no caso des-

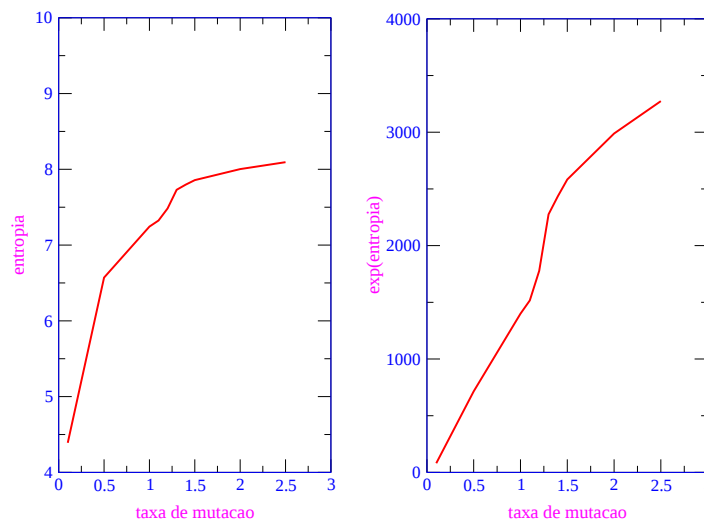


Figura 6.9: Resultados da simulação para organismos digitais. Temos na parte a) Entropia e (b) exponencial da entropia.

sas apresentarem uma maior neutralidade no seu relevo local. Em resumo, um organismo que se replica mais rapidamente e que ocupa um pico estreito no relevo de adaptação - onde há poucos mutantes adjacentes com valores seletivos comparáveis - pode ser eliminado por um organismo que ocupa um pico menor porém mais plano, onde muitos mutantes adjacentes têm valores seletivos similares. Em essência, "sobrevivência do mais suave" (*survival of the flattest*) pode ser tão importante quanto "sobrevivência do mais apto" (*survival of the fittest*) para taxas de mutações elevadas [82].

Podemos visualizar este problema como uma transição de um comportamento não-altruísta para um regime altruísta no contexto de *seleção de grupo*. A região em que o organismo mais apto domina a população pode ser vislumbrada como um regime não-altruísta, desde que este só gera, por meio de mutação, indivíduos bem menos aptos e, portanto, não colabora para a manutenção e coexistência de um grupo de genótipos. Já para o segundo regime, onde a quase-espécie situada no pico mais suave domina a população, podemos associá-lo a um regime altruísta, desde que neste regime um grupo é promovido. O grupo de mutantes que coexistem nesse regime estão inter-conectados por

meio do processo de mutação. Como esse regime ocorre para grandes valores de taxas de mutação, o indivíduo quando se replica muito provavelmente produzirá um indivíduo diferente de si, o que pode ser encarado como um custo para sua auto-replicação, entretanto é esse mecanismo que garante a promoção do grupo ao qual pertence.

6.2 O modelo NK

6.2.1 Descrição do modelo

O modelo NK é uma espécie de versão genética de um modelo físico aplicado a vidros de spin (que são um tipo de material magnético desordenado), o *modelo de vidros de spin com interação de multispins*, introduzido por Bernard Derrida [83–85]. A associação direta entre a abordagem utilizada no estudo de vidros de spin e da teoria evolucionária tem sido cada vez mais constante [26, 86]. A grande virtude do modelo NK é mostrar como aspectos distintos do genótipo resultam em variações das propriedades elementares dos relevos de adaptação, permitindo desta forma o estudo de uma família de *relevos de adaptação* de forma controlada.

Vamos agora a uma breve descrição do modelo. Consideremos um organismo composto por N segmentos, ou genes, cada qual assumindo dois estados alternativos: 0 e 1, ou seja, estamos considerando seqüências binárias de comprimento N . Um organismo é uma combinação única dos estados 1 ou 0 de cada um destes N segmentos. Desta forma, o espaço de genótipos é constituído de 2^N combinações possíveis de seqüências, cada uma definindo um genoma distinto. Assumimos que o valor adaptativo de cada um dos indivíduos presentes na população depende unicamente da composição de seu genoma. Em geral, a contribuição de cada um dos N genes para o valor seletivo total do genoma é uma função bastante complexa dos estados de vários outros genes formando o indivíduo. Os geneticistas denominam esta dependência, ou acoplamento entre os genes, de *epistase ou interação epistática*. Desta forma, genes situados em regiões distintas do

mesmo genótipo afetam a contribuição de um dado gene para a adaptação do organismo. Em particular, assumimos que a *epistase* é modelada atribuindo-se a cada gene K outros elementos, dentre os $N - 1$ restantes, que determinarão o valor seletivo local. O valor seletivo $F(\mathbf{x})$ de uma configuração de genótipo $\mathbf{x} = (x_1, x_2, \dots, x_N)$ é então definida como

$$F(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N f_i, \quad (6.3)$$

onde a contribuição do sítio i , f_i , é uma função aleatória dependendo de x_i e outros K x_l 's. Mais precisamente, f_i é uma função aleatória de 2^{K+1} argumentos com valores uniformemente distribuídos no intervalo $(0, 1]$. Na Figura 6.10 exibimos um exemplo ilustrativo para o caso $N = 4$ e $K = 2$. Este tipo de formulação para o fenômeno da epistases é bastante semelhante à formulação apresentada para o estudo de redes regulatórias de genes, onde a ativação ou desativação de um dado gene na rede depende dos sinais enviados por outros genes, e desta forma determinam o seu padrão de comportamento [87].

O modelo NK é fascinante devido a um ponto essencial: alterando o número de genes nas interações epistáticas por gene, K , podemos alterar o número de picos existentes no relevo. Quanto maior for K - quanto mais inter-conectados estiverem os genes - mais vínculos conflitantes existirão, de forma que o relevo de adaptação torna-se mais rugoso, ou seja, possui um número maior de máximos locais. Desta forma, o parâmetro K controla a rugosidade do *relevo de adaptação*. Como dois exemplos extremos podemos citar o caso $K = 0$, no qual a mudança no valor seletivo devido à mudança de um único elemento x_i (mutação pontual) é da ordem de $1/N$, e o relevo neste caso é considerado suave. Por outro lado, para $K = N - 1$ uma única mutação pontual altera todos os f_i 's, e neste caso o relevo torna-se extremamente *rugoso*. Casos intermediários correspondem a valores intermediários de K .

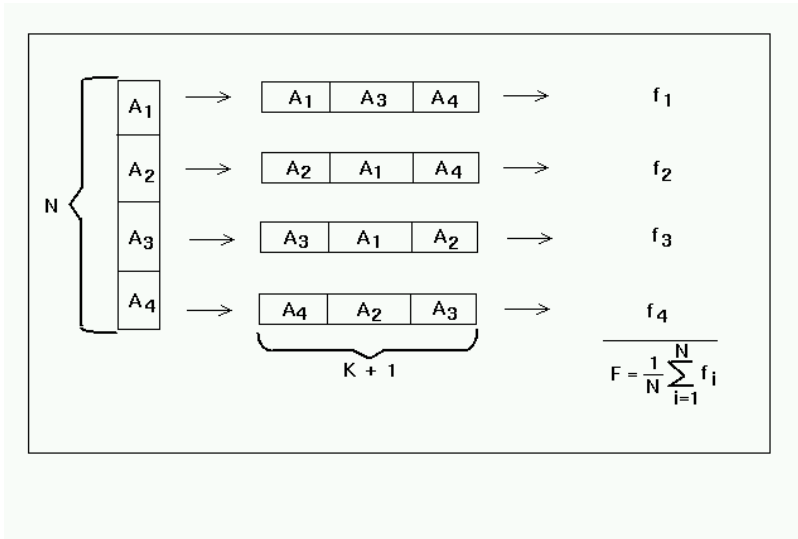


Figura 6.10: Cálculo do valor adaptativo F no modelo NK para o caso $N = 4$ e $K = 2$. $N = 4$ genes compõem o genoma cada qual com um alelo A_i que pode assumir os valores 0 ou 1. Para cada um dos 4 alelos um subgenoma é construído tomando o próprio gene e $K = 2$ outros genes, aleatoriamente escolhidos entre os 3 genes restantes no genoma. A cada subgenoma é atribuído um valor f_i tomado aleatoriamente do intervalo $(0, 1]$. O valor seletivo total F para o genoma é então igual a média dos $N = 4$ valores dos subgenomas.

6.2.2 Catástrofe de erro

Aqui apresentamos alguns resultados que demonstram que o fenômeno da catástrofe de erro não é um evento exclusivo para o relevo de pico único. Na verdade, está presente em relevos reais de adaptação como também em outros relevos oriundos de modelos teóricos [88–90].

Em nossa dinâmica estocástica aqui utilizada, seqüências binárias são otimizadas com respeito às suas propriedades cinéticas por meio de um processo estocástico envolvendo seleção e mutação. No processo de reprodução uma fração da população é escolhida para produzir descendentes. Eles são escolhidos por meio de seleção atribuindo maior probabilidade de reprodução àqueles indivíduos com maior valor de adaptação. Assumimos também que o processo de replicação está sujeito à falhas, ou seja, a seqüência clonada pode sofrer mutações. Supomos uma probabilidade de mutação por dígito, u , uniforme ao longo da seqüência.

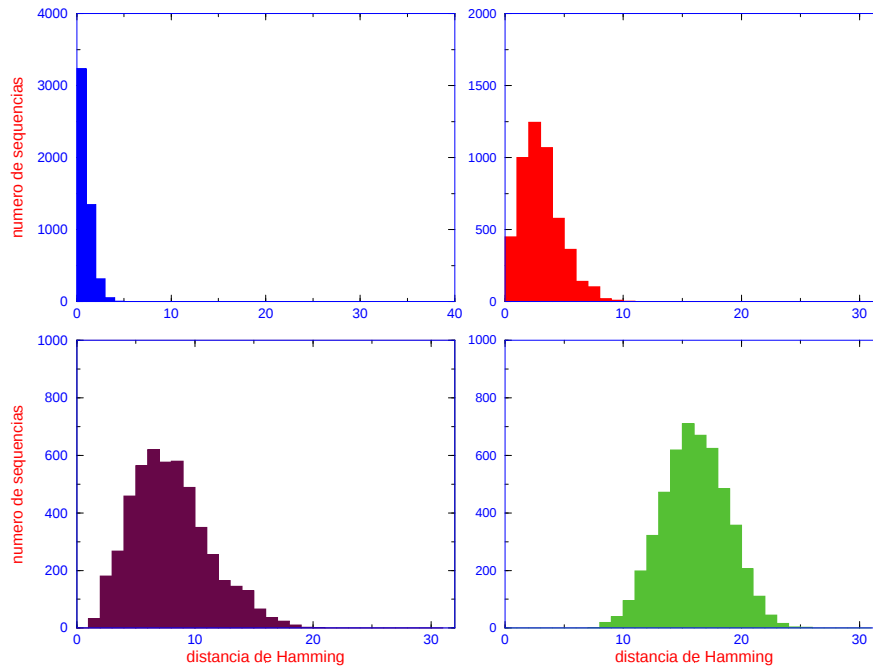


Figura 6.11: Histograma da distribuição da distância de Hamming entre as seqüências e a seqüência consenso da população. Para esta simulação temos população de tamanho $N = 5000$, seqüências de tamanho $L = 32$, e parâmetro $K = 8$. A probabilidade de uma mutação pontual é igual a: (a) $u = 0.001$, (b) $u = 0.005$, (c) $u = 0.01$ e (d) $u = 0.1$

Na Figura 6.11 podemos visualizar a distribuição da distância de Hamming entre as seqüências e a seqüência consenso na população. A seqüência consenso é definida como a seqüência dos símbolos que são predominantes em cada posição. Em nossa simulação utilizamos uma população de tamanho médio igual a $N = 5000$, seqüências de comprimento $L = 32$ e valor de parâmetro $K = 8$. O número de gerações para a realização da medida é igual a $t = 10000$. Nessa Figura exibimos a distribuição de mutantes para quatro valores distintos de u .

Na Figura 6.12 exibimos a evolução temporal do valor adaptativo médio na população para os casos mostrados na Figura 6.11. Dessa figura podemos observar claramente três regimes distintos. Na parte (a) da figura, onde $u = 0.001$, visualizamos que a grande maioria das seqüências compondo a população são idênticas à seqüência consenso, sendo esta também a dominante. Na parte (b), para valor de $u = 0.005$, observamos que mais

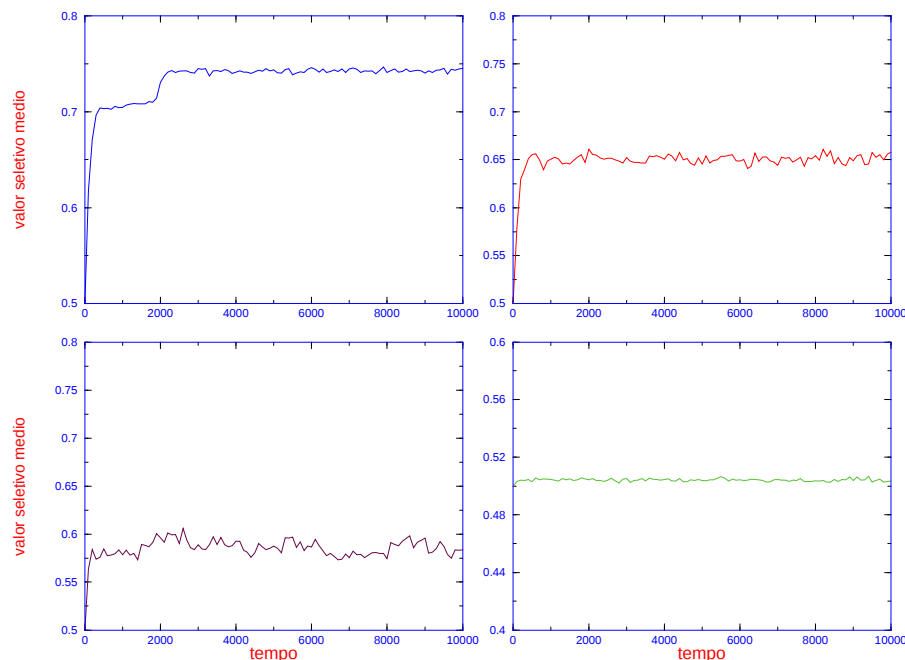


Figura 6.12: Evolução do valor adaptativo médio da população. Os parâmetros são os mesmos daqueles utilizados na Figura 6.11. A probabilidade de uma mutação pontual é igual a: (a) $u = 0.001$, (b) $u = 0.005$, (c) $u = 0.01$ e (d) $u = 0.1$

uma vez a maioria das seqüências são iguais ou muito próximas à seqüência consenso, e a largura da distribuição é relativamente pequena. Neste regime, a informação genética é conservada e a seqüência consenso pode ser transmitida para gerações futuras.

Já na parte (c) da mesma figura, observamos um outro regime, onde os elementos da população na grande maioria diferem da seqüência consenso. Neste regime, a distribuição de mutantes em torno da seqüência consenso desenvolve uma caminhada evolutiva pelo espaço de genótipos. Podemos observar também que este regime é caracterizado por um valor seletivo médio superior ao valor seletivo médio do espaço de genótipos, que é igual a $1/2$ (ver Figura 6.12). Uma clara evidência de que a distribuição dos elementos na população possui uma certa estrutura ou ordenamento.

Na parte (d), onde $u = 0.1$, obtemos uma distribuição completamente desordenada das seqüências no espaço de genótipos, ou seja, estas disseminam-se completamente e a probabilidade de encontrar uma seqüência em qualquer sítio nesse espaço obedece uma

distribuição uniforme, ou seja, iguala a $1/2^L$. Portanto, a concentração de mutantes diferindo i dígitos de uma dada sequência consenso, no limite $N \rightarrow \infty$, é igual a $Y_i = \binom{L}{i} \frac{1}{2^L}$. Para uma população finita o número de mutantes de uma dada classe i , é aproximadamente dado por $N_i \approx \binom{L}{i} \frac{N}{2^L}$. Neste caso, temos comportamento idêntico ao observado no regime estocástico do modelo de quase-espécie para a função seletiva de pico único. Através da análise da Figura 6.12 (d) vemos que o valor seletivo médio da população é praticamente igual a $1/2$. Indicando uma dispersão total da população no espaço de genótipos.

O limiar de erro para o relevo aqui discutido corresponde ao ponto onde o segundo regime começa a existir. Neste ponto a informação genética é perdida desde que uma sequência consenso estável não pode ser mantida na população, apesar da existência de um certo ordenamento na distribuição dos mutantes.

Os resultados e comportamentos aqui verificados estendem-se para outros valores de parâmetro K .

6.2.3 Velocidade de Adaptação

Na Figura 6.13 exibimos exemplos típicos para a evolução temporal do valor adaptativo de uma população. Mostramos o seu comportamento para vários valores de parâmetro K e de taxa de mutação u .

Dessa figura podemos verificar que a mutação tem papel preponderante no processo de otimização evolucionária. Em particular, o aspecto mais interesse é a observação da velocidade de adaptação do sistema com a taxa de mutação. Podemos identificar de forma precisa que, a medida que incrementamos o valor da probabilidade de mutação u , existe um aumento dessa velocidade. Ou seja, cada vez mais rapidamente as sequências conseguem encontrar máximos locais dentro do relevo de adaptação de maior valor seletivo, e de forma mais eficiente. Entretanto, a partir de um certo valor crítico, esse

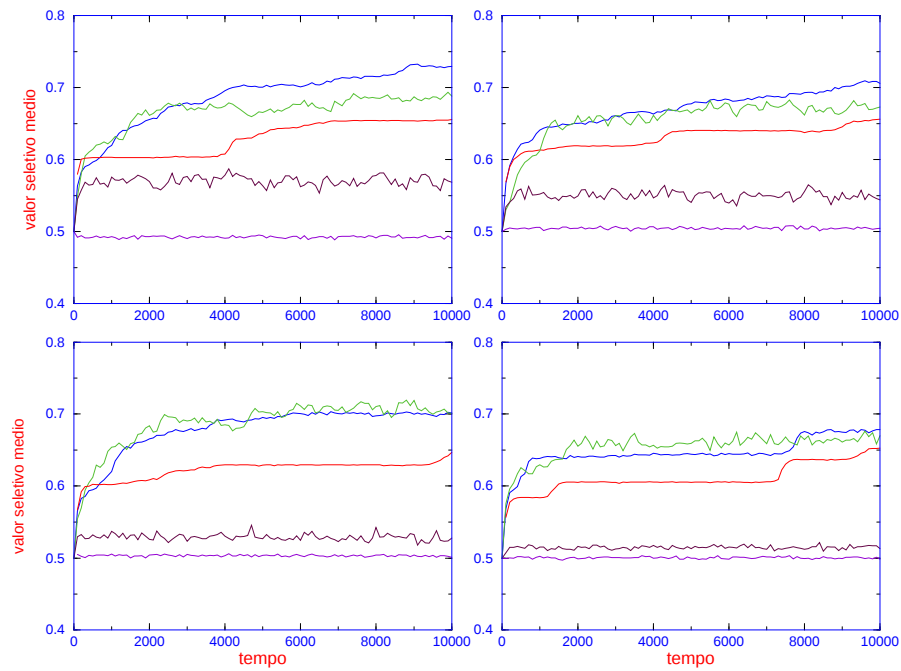


Figura 6.13: Evolução temporal do valor adaptativo médio da população. Consideramos seqüências de tamanho $L = 64$. Os valores de K utilizados na simulação são: (a) $K = 2$, (d) $K = 4$, (c) $K = 8$ e (d) $K = 16$. As cores denotam os diferentes valores da probabilidade de mutação por dígito, e obedecem a seguinte correspondência: $u = 0.00001$ (vermelho), $u = 0.0001$ (azul), $u = 0.001$ (verde), $u = 0.01$ (marrom escuro), $u = 0.1$ (violeta).

comportamento se altera, e um maior incremento de u leva então a uma redução da eficiência do processo. A razão para isto, deve-se ao fato de que com o incremento de u também reduzimos a possibilidade de fixação da população nesses ótimos locais. Um aumento ainda maior do valor de mutação leva a um total desordenamento da população no espaço de genótipos.

Este é um resultado interessante por salientar claramente a importância da interação entre seleção natural e mutação. O papel primordial da mutação é criar uma diversidade de elementos de forma a proporcionar uma atuação contínua e precisa da seleção natural, resultando em uma otimização constante do sistema.

Capítulo 7

Conclusão

O uso de técnicas desenvolvidas na Mecânica Estatística para investigar problemas em outras áreas da ciência tem sido cada vez mais constante. Problemas com motivação biológica, econômica e mesmo sociais têm atraído a atenção dos físicos, e isto tem colaborado para o desenvolvimento dessas áreas como também para o aprimoramento das técnicas desenvolvidas. O trabalho aqui realizado faz uso de diversas técnicas aplicadas em problemas em Física em geral. Como exemplo, citamos o uso da técnica de escalonamento de tamanho finito, sistemas dinâmicos, dinâmica estocástica e métodos de simulação computacional. O tema aqui proposto possui grande motivação biológica e para as teorias de origem da vida. Mais precisamente, abordamos o problema da emergência de organização e informação em sistemas pré-bióticos.

O estudo de modelos e mecanismos que garantam a coexistência de elementos distintos e a manutenção da informação genética no processo evolucionário é de importância fundamental para a compreensão da evolução dos organismos modernos a partir do cenário pré-biótico. Nesse contexto, o trabalho de Manfred Eigen foi um dos precursores na descrição matemática de sistemas de moléculas auto-replicas sujeitos aos processos de seleção natural e mutação. O modelo de quase-espécies de Eigen apresenta resultados bastante intrigantes para as teorias de origem da vida. O mais relevante deles é a observação de um limiar crítico para a probabilidade de replicação exata de uma macromolécula,

a partir da qual toda a informação genética relevante é perdida. Esse fenômeno, conhecido como *limiar* ou *catástrofe de erro*, impõe um limite máximo para o tamanho L ou complexidade máxima das moléculas auto-replicas.

A caracterização completa da transição de limiar de erro para valores finitos de L ainda não havia sido realizada. De fato, similarmente à definição de temperatura crítica para redes finitas, não há uma definição consensual para o limiar de erro para valores finitos de L . Contudo, o estudo dos desvios sistemáticos do comportamento crítico no limite de $L \rightarrow \infty$ introduzidos por efeitos de tamanho finito, que são praticamente independentes da definição adotada, nos dá informação relevante sobre o comportamento das propriedades macroscópicas próximo à região crítica [37]. Assim, nesta tese realizamos de forma pioneira o estudo dos efeitos de tamanho finito das moléculas próximo à transição de limiar de erro. Em particular, determinamos a largura da transição, ou seja, o intervalo de Q em torno de Q_c onde as características da transição persistem, onde $Q_c = 1/a$ é o ponto de transição no limite de $L \rightarrow \infty$. Como esperado encontramos que a largura da região de transição tende a zero de acordo com $L^{-\frac{1}{\nu}}$ e, por meio do procedimento padrão do colapso de dados [37], encontramos que $\nu = 1$, independentemente dos valores da vantagem seletiva a . Analogamente à magnetização e susceptibilidade magnética definidos para o problema de sistemas de spins, introduzimos a distância de Hamming normalizada entre a sequência mestra e a população como parâmetro de ordem do modelo, dada por $d = \frac{1}{L} \sum_{P=0}^L P Y_P$, e $\sigma^2 = L^2 \sum_{P=0}^L \left(\frac{P}{L} - d\right)^2 Y_P$ é o desvio quadrático médio. Verificamos que tanto d como σ^2 são bem descritas por meio das equações $d(L, Q) = \tilde{d}((Q - Q_c)L^{1/\nu})$ e $\sigma^2(L, Q) = L^{\gamma/\nu} \tilde{\sigma}^2((Q - Q_c)L^{1/\nu})$, onde $\tilde{d}$ e $\tilde{\sigma}^2$ são funções de escala, ou seja, independem de L . Estimamos que a razão dos expoentes críticos γ/ν é igual a 1.96 independentemente do valor de a , estando este resultado em ótima concordância com o resultado analítico obtido por Galluccio ($\nu = 1$ e $\gamma = 2$) por meio do mapeamento do modelo em um problema de polímeros direcionados em um meio aleatório [35]. Da mesma forma que

no caso de problemas combinatoriais [2], verificamos que o uso da técnica de escala de tamanho finito é bastante efetivo para a caracterização da transição de limiar de erro no modelo de quase-espécies.

Outra situação de interesse em que a transição do limiar de erro não é bem definida é o regime de população finita (isto é, N finito) mas o tamanho das seqüências é infinito. Essa versão estocástica do modelo de quase-espécies é estudada em detalhes no capítulo 3. Neste caso, consideramos apenas duas classes distintas de seqüências: a primeira consistindo apenas de seqüências mestras, de valor seletivo a ; e a segunda consistindo de todos os demais tipos, cujo valor seletivo é igual a 1. Como em nossa formulação consideramos seqüências de tamanho infinito, $L \rightarrow \infty$, podemos desprezar as mutações favorecendo a classe de seqüências mestras (mutações reversas). O processo dinâmico é descrito por uma cadeia de Markov, onde cada estado possível do sistema é caracterizado pela variável $n = 0, \dots, N$, que dá o número de seqüências mestras na população. A probabilidade de transição entre estados em gerações consecutivas é dada pela matriz de transição T de acordo com a Equação (3.6). A dinâmica considerada consiste de dois passos independentes: no primeiro passo a seleção natural atua, onde atribuímos maior chance de reprodução àqueles indivíduos mais aptos. No segundo passo, os indivíduos que irão compor a geração futura são submetidas ao processo de mutação. A evolução temporal do vetor distribuição de probabilidades $\mathbf{P}(t)$, onde cada elemento $P_n(t)$ representa a probabilidade de encontramos n seqüências mestras no instante t na população, é obtido através da relação de recorrência $\mathbf{P}(t) = \mathbf{T}^t \mathbf{P}(0)$. Escrevendo $P(0)$ como uma combinação linear dos autovetores de T , derivamos uma expressão para o elemento $P_0(t)$, que descreve a probabilidade de termos $n = 0$ seqüências mestras na população como função dos autovalores da matriz de transição. Para tempos muito grandes ($t \rightarrow \infty$) o comportamento de $P_0(t)$ é inteiramente governado pelo segundo autovalor dominante da matriz (λ_1). No regime assintótico, temos sempre como solução $P_0(t) = 1$. O tempo de

relaxação, definido como o tempo necessário para que o número de mestras se reduza a $1/e$ de seu valor inicial é definido por $\tau \equiv 1/\lambda_1$. Verificamos que esse tempo característico é bem descrito por $\tau = N^{1/2} f_a [(Q - Q_c) N^{1/2}]$, onde f_a é uma função de escala dependente de a mas independente de N . A transição em nosso modelo é constatada através da divergência de τ em $Q_c = 1/a$ no limite $N \rightarrow \infty$. Sendo que a largura da transição decresce de acordo com $N^{-1/2}$. A grande vantagem de nossa abordagem é que não utilizamos nenhuma definição arbitrária para o limiar de erro para populações finitas.

Como alternativa ao modelo de quase-espécies e na tentativa de solucionar a crise de informação associada ao limiar de erro, o próprio Eigen [4, 5] introduziu o modelo de hiperciclos. Os hiperciclos são cadeias catalíticas fechadas, ou ciclos catalíticos, onde cada elemento auxilia, por meio de catálise, a replicação do elemento imediatamente posterior na organização cíclica. Embora diversos estudos tenham sido desenvolvidos, a grande maioria lida somente com aspectos da coexistência dos elementos do hiperciclo. No capítulo 4, abordamos a questão quase que inexplorada da propagação de erro nos hiperciclos. Em particular, estudamos analiticamente o regime de estado estacionário de uma rede com n tipos distintos de seqüências, formando um hiperciclo assimétrico, e sua cauda de erro. O hiperciclo é dito assimétrico pois assumimos a existência de uma seqüência mestra que possui uma vantagem seletiva a com relação aos outros elementos. Estudamos a dinâmica de hiperciclos de vários tamanhos, obtendo seus pontos fixos e realizando o estudo padrão de estabilidade. De posse desses dados, construímos o diagrama de fases no espaço de parâmetros do modelo, descrevendo a região de existência e estabilidade das soluções de estado estacionário. Diferentemente dos resultados para os hiperciclos simétricos, no caso assimétrico cadeias catalíticas são estáveis dentro de regiões delimitadas no espaço de parâmetros. O resultado mais interessante de sua investigação é a observação de um limiar de erro (Q_h) para hiperciclos com $n \leq 4$, que torna-se menor que aquele verificado no modelo de quase-espécies ($Q_1 = 1/a$) para valores de constante catalíticas da ordem

de a^2 (veja Eq. (4.63)).

Os modelos de compartimentalização têm sido estudados no contexto pré-biótico devido às conjecturas para o cenário evolutivo naquele período. No Capítulo 5, estudamos a evolução da produção de enzimas e os efeitos de sinergia utilizando idéias de seleção de grupo para a evolução do comportamento altruísta. Usamos a formulação de Wilson para o modelo de *demes* estruturados, porém aqui, os valores seletivos dos indivíduos dependem unicamente da composição interna da deme da qual fazem parte. Para ambos os modelos investigados, obtemos os diagramas de fases delimitando as regiões de estabilidade dos pontos fixos. Interessantemente, obtemos regiões delimitando um regime de coexistência entre os elementos altruístas e não-altruístas dentro de um mesmo grupo, resultado não obtido da análise de outros modelos considerados [60–62]. Verificamos que a introdução de mutação no sistema leva a mudanças qualitativas no estado estacionário. Em particular, verificamos um limiar de erro a partir do qual não é mais possível realizar uma distinção entre os regimes altruístas e não-altruístas.

Por fim, no capítulo 6 apresentamos alguns resultados preliminares para relevos de replicação muito mais complexos do que os analisados anteriormente, dando ênfase ao papel da neutralidade e rugosidade dos relevos na evolução do sistema. Por meio de um modelo teórico simples, mostramos que mesmo para valores seletivos superiores, uma população pode ser excluída por outra de menor valor seletivo dependendo do grau de neutralidade da região do relevo em que se encontram e também da taxa de mutação das seqüências. Por meio de alguns resultados obtidos na plataforma de vida artificial Avida [78] mostramos o mesmo tipo de comportamento verificado na modelagem teórica. Descrevemos também o modelo NK de Kauffman que gera uma família de relevos rugosos cuja rugosidade pode ser controlada variando-se certos parâmetros de controle. Mostramos que mesmo nesse tipo de relevo o fenômeno da catástrofe de erro é também observado. Entretanto, observamos três regimes distintos de comportamento. O primeiro é caracte-

rizado por um aglomerado ordenado de seqüências em torno da seqüência de consenso. A partir de um certo valor de mutação essa seqüência consenso desloca-se constantemente pelo espaço de genótipos juntamente com uma distribuição de mutantes. Neste segundo regime o sistema não é mais capaz de passar para gerações futuras a informação sobre a seqüência consenso na população. No terceiro regime, que ocorre para valores elevados de mutação, existe um desordenamento total e a população é completamente espalhada pelo espaço de genótipos.

Como perspectivas futuras de trabalho, desejamos dar continuidade ao trabalho apresentado no capítulo 6. Em particular, queremos investigar as propriedades da dinâmica evolucionária em relevos rugosos em geral, incluindo os relevos derivados de sistemas biológicos, como também aqueles obtidos da teoria de vidros de spins. De grande interesse é a determinação dos diferentes regimes de comportamento do sistema com a taxa de mutação das seqüências. O resultado interessante de que a taxa de mutação determina a performance do processo de otimização no modelo NK, nos estimula a analisarmos sob que condições topológicas de relevos em geral esse resultado é sempre verificado. Sugerimos ainda, um estudo com relação à introdução da neutralidade nesses relevos e suas conseqüências no processo de otimização de uma população.

Apêndice A

Solução Analítica das Eqs. do modelo de quase-espécies

No capítulo 2 apresentamos a seguinte forma para as equações diferenciais descrevendo a evolução temporal das frequências relativas dos diversos tipos de indivíduos na população:

$$\dot{x}_i(t) = [W_{ii} - \bar{E}(t)] x_i(t) + \sum_{k \neq i} W_{ik} x_k(t). \quad (\text{A.1})$$

Devido ao fato que o termo $\bar{E}(t)$ é inerentemente não-linear, essas equações podem ser resolvidas por meio de aproximações correspondentes à teoria de perturbação de segunda ordem.

Soluções exatas também podem ser derivadas para estas equações [91, 92]. Nós seguimos aqui essencialmente o método de Jones et al. [92].

O termo não-linear $\bar{E}(t)$ pode ser removido por meio da seguinte transformação de variáveis:

$$x_i(t) = z_i(t) f_t \quad \text{com} \quad f(t) = \exp \left(\int_0^t \bar{E}(\tau) d\tau \right), \quad i = 1, 2, \dots, 2^L, \quad (\text{A.2})$$

resultando no seguinte conjunto de equações diferenciais para as coordenadas $z(t)$:

$$\dot{z}_i(t) = \sum_k W_{ik} z_k(t), \quad (\text{A.3})$$

onde a soma agora inclui também os termos diagonais. Percebemos da Equação (A.1)

que, com o vínculo $\sum x_k(t) = 1$, $f(t)$ pode ser escrito como

$$f(t) = \left(\sum_k z_k(t) \right)^{-1}. \quad (\text{A.4})$$

A solução das Eqs. (A.3) seguem o procedimento padrão de álgebra linear.

Consideremos cada variável $z_i(t)$ como o i -ésimo componente de um vetor coluna $\mathbf{z}$ e cada coeficiente de taxa W_{ik} como um elemento da matriz quadrática $\mathbf{W}$. O sistema de Equações (A.3) pode ser então escrito como

$$\frac{d\mathbf{z}(t)}{dt} = \mathbf{W}\mathbf{z}(t). \quad (\text{A.5})$$

Dado que a matriz $\mathbf{W}$ é diagonalizável, o sistema de equações diferenciais pode ser desacoplado por meio da transformação

$$\mathbf{z}(t) = \mathbf{L}\zeta(t) \quad e \quad \zeta(t) = \mathbf{L}^{-1}\mathbf{z}(t), \quad (\text{A.6})$$

onde $\zeta(t)$ representa novamente um vetor coluna, e $\mathbf{L}$, como também sua inversa são matrizes quadráticas. A combinação das eqs (A.5) e (A.6) nos dá

$$\frac{d\zeta}{dt} = \mathbf{L}^{-1}\mathbf{W}\mathbf{L}\zeta(t). \quad (\text{A.7})$$

A matriz $\Lambda = \mathbf{L}^{-1}\mathbf{W}\mathbf{L}$ é diagonal como suposto anteriormente. Seus elementos são autovalores de W : λ_k ($k = 1, 2, \dots$). Os autovetores correspondentes $\mathbf{l}_k = \{l_{1k}, l_{2k}, \dots\}$ são as composições estacionárias para $\zeta(t)$, que crescem em magnitude a uma taxa λ_k . Desde que os elementos não-diagonais da matriz $\mathbf{W}$ são estritamente positivos, o Teorema de Frobenius se aplica. Então, o maior autovalor $\lambda_{max} = \lambda_0$ é não-degenerado. Sem perda de generalidade, ordenamos os autovalores

$$\lambda_0 > \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \geq \dots \quad (\text{A.8})$$

O autovetor correspondente a λ_0 é denominado de autovetor dominante $\mathbf{l}_0$. Todas as suas componentes são estritamente positivos: $l_{i0} > 0$ ($i = 1, 2, \dots$). Os autovetores a esquerda são dados pelas colunas da matriz inversa $\mathbf{L}^{-1}$.

Para cada componente de $\zeta(t)$ obtemos uma solução da forma

$$\zeta_k(t) = \zeta_k(0) \exp(\lambda_k t). \quad (\text{A.9})$$

Os autovalores λ_k da matriz $\mathbf{W}$ são soluções para λ da equação de determinante

$$\det(\mathbf{W} - \lambda \mathbf{I}) = 0, \quad (\text{A.10})$$

onde $\det$ denota o determinante e $\mathbf{I}$ a matriz unitária. A avaliação da Equação (A.10) resulta em um polinômio cujo grau é dado pelo número de variáveis de concentração. De forma a determinar os componentes dos autovetores a esquerda e a direita da matriz $\mathbf{W}$, iniciamos de

$$\mathbf{W}\mathbf{L} = \mathbf{L}\mathbf{\Lambda} \quad (\text{A.11})$$

e

$$\mathbf{L}^{-1}\mathbf{W} = \lambda \mathbf{L}^{-1} \quad \text{ou} \quad \mathbf{W}'(\mathbf{L}^{-1})' = (\mathbf{L}^{-1})'\mathbf{\Lambda}. \quad (\text{A.12})$$

A linha denota a matriz transposta. Explicitamente em termos dos elementos de matriz, $\mathbf{L} = \{l_{ij}\}$ e $\mathbf{L}^{-1} = \{h_{ij}\}$, as duas equações são escritas como

$$\sum_j W_{ij} l_{jk} = l_{ik} \lambda_k, \quad (\text{A.13})$$

$$\sum_j W_{ji} h_{kj} = h_{ki} \lambda_k. \quad (\text{A.14})$$

As soluções explícitas $z_i(t)$ da Equação (A.3) são obtidas como superposições dos modos normais individuais,

$$z_i(t) = \sum_k l_{ik} c_k \exp(\lambda_k t), \quad i = 1, 2, \dots \quad (\text{A.15})$$

As constantes c_k podem ser obtidos das condições iniciais, $z_i(0) = x_i(0)$:

$$c_k = \sum_i h_{ki} x_i(0). \quad (\text{A.16})$$

As soluções explícitas para as variáveis de concentração relativa $x_i(t)$ podem ser obtidas das Equações (A.2) e (A.4):

$$x_i(t) = \frac{\sum_k l_{ik} c_k \exp(\lambda_k t)}{\sum_j \sum_k l_{jk} c_k \exp(\lambda_k t)}. \quad (\text{A.17})$$

A distribuição de seqüências então converge assintoticamente para o autovetor dominante, $\mathbf{l}_0$ da matriz W .

Apêndice B

Cálculo dos elementos da matriz M de mutação entre classes

Os elementos M_{PR} descrevem a probabilidade de mutação de uma molécula de classe R para uma molécula de classe P . Aqui P e R correspondem ao número de 1's que compõem a seqüência, conforme nossa definição de classes. Lembremos que q é a probabilidade de replicação exata por dígito. Vamos dividir a seqüência em consideração em dois grupos: um composto pelos R 1's e o outro pelos $(L - R)$ 0's. primeiramente vamos encontrar uma expressão para a probabilidade de mutação dos dígitos do primeiro grupo. Supondo que dos R elementos deste grupo I sejam copiados corretamente e o restante sofra mutação, temos portanto

$$P_{IR} = \binom{R}{I} q^I (1 - q)^{R-I} \quad (\text{B.1})$$

que nos dá a probabilidade desta situação ocorrer. Consideremos agora os dígitos pertencentes ao segundo grupo. A probabilidade de que I' dos seus elementos $(L - R)$ sejam copiados erroneamente (isto é, mutem para 1's) é simplesmente

$$P_{I'(L-R)} = \binom{L-R}{I'} q^{L-R-I'} (1 - q)^{I'}. \quad (\text{B.2})$$

Afim de que a seqüência resultante possua P 1's, devemos ter

$$I' + I = P. \quad (\text{B.3})$$

Ambos os processos acima descritos ocorrem simultaneamente e de forma independente, logo

$$M_{PR} = \sum_{I=I_l}^{I_u} \binom{R}{I} \binom{L-R}{P-I} q^{L-P-R+2I} (1-q)^{P+R-2I}, \quad (\text{B.4})$$

com $I_l = \max(0, P+R-L)$ e $I_u = \min(P, R)$.

Apêndice C

Método das potências ou iteração vetorial

C.1 O autovalor dominante e o autovetor associado de uma matriz

Aqui consideramos o problema de autovalores para uma matriz diagonalizável e real $\mathbf{A}$: onde $\mathbf{A}$ possui n autovetores linearmente independentes $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbf{R}^n$. Considerando um vetor inicial arbitrário $\mathbf{z}^0 \neq \mathbf{0}$ em $\mathbf{R}^n$, podemos realizar uma seqüência de iterações $\{\mathbf{z}^{(\nu)}\}$ de acordo com a regra

$$\mathbf{z}^{(\nu)} := \mathbf{A}\mathbf{z}^{(\nu-1)} \quad \text{onde} \quad \mathbf{z}^{(\nu-1)} = \begin{pmatrix} z_1^{(\nu-1)} \\ z_2^{(\nu-1)} \\ \vdots \\ z_n^{(\nu-1)} \end{pmatrix} \quad \text{para} \quad \nu = 1, 2, \dots \quad (\text{C.1})$$

Supondo que os autovetores $\mathbf{x}_i$ de $\mathbf{A}$ formam a base para $\mathbf{R}^n$, $\mathbf{z}^0$ pode ser escrito unicamente como uma combinação linear dos $\mathbf{x}_i$:

$$\mathbf{z}^0 = \sum_{i=1}^n c_i \mathbf{x}_i, \quad (\text{C.2})$$

onde $c_i \neq 0$ para no mínimo um valor de i . Assim, como $\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i$ para cada i , obtemos

$$\mathbf{z}^{(\nu)} = c_1 \lambda_1^\nu \mathbf{x}_1 + c_2 \lambda_2^\nu \mathbf{x}_2 + \dots + c_n \lambda_n^\nu \mathbf{x}_n \quad \forall \quad \nu = 0, 1, 2, \dots \quad (\text{C.3})$$

No próximo passo queremos estudar os quocientes $q_i^{(\nu)}$ entre i -ésimos componentes de $\mathbf{z}$ obtidos em iterações sucessivas $\mathbf{z}^{(\nu)}$ e $\mathbf{z}^{(\nu-1)}$:

$$q_i^{(\nu)} := \frac{z_i^{(\nu)}}{z_i^{(\nu-1)}} = \frac{c_1 \lambda_1^\nu \mathbf{x}_{1,i} + c_2 \lambda_2^\nu \mathbf{x}_{2,i} + \dots + c_n \lambda_n^\nu \mathbf{x}_{n,i}}{c_1 \lambda_1^{\nu-1} \mathbf{x}_{1,i} + c_2 \lambda_2^{\nu-1} \mathbf{x}_{2,i} + \dots + c_n \lambda_n^{\nu-1} \mathbf{x}_{n,i}}. \quad (\text{C.4})$$

Caso)

$\mathbf{A}$ tem um único autovalor dominante, $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \dots \geq |\lambda_n|$:

a) Se $c_1 \neq 0$ e $x_{1,i} \neq 0$ para algum i , então

$$q_i^{(\nu)} = \lambda_1 + O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^\nu\right), \quad \text{ou seja,} \quad \lim_{\nu \rightarrow \infty} q_i^{(\nu)} = \lambda_1. \quad (\text{C.5})$$

Desde que $x_{1,i} \neq 0$ para no mínimo um i , desta forma pelo menos um dos quocientes $q_i^{(\nu)}$ tende a λ_1 .

Para valores de ν suficientemente grandes, $q_i^{(\nu)}$ será uma aproximação para o maior autovalor λ_1 de $\mathbf{A}$, denominado autovalor dominante de $\mathbf{A}$. Assim,

$$q_i^{(\nu)} \approx \lambda_1. \quad (\text{C.6})$$

Na prática, iteramos a Eq. (C.1) até que Eq. (C.6) seja uniformemente válida para todos os índices i com $x_{1,i} \neq 0$. Para grandes valores de ν , o vetor $\mathbf{z}^{(\nu)}$ é aproximadamente paralelo a $\mathbf{x}_1$. Para $\nu \rightarrow \infty$ temos

$$\mathbf{z}^{(\nu)} \sim \lambda_1^\nu c_1 \mathbf{x}_1 \quad \text{e} \quad \mathbf{z}^{(\nu)} \sim \lambda_1 \mathbf{z}^{(\nu-1)}. \quad (\text{C.7})$$

Se os autovetores $\mathbf{x}_i$ têm comprimento unitário, então assintoticamente

$$\frac{\mathbf{z}^{(\nu)}}{|\mathbf{z}^{(\nu)}|} \approx \mathbf{x}_1. \quad (\text{C.8})$$

C.1.1 Cálculo do segundo autovalor

Seja λ_1 o autovalor de maior valor absoluto da matriz $\mathbf{A}$ e, $\mathbf{x}_1$ o seu autovetor correspondente. Supomos que seja possível tornar a primeira componente de $\mathbf{x}_1$ igual a 1, ou

seja,

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ u_{12} \\ u_{13} \\ \vdots \\ u_{1n} \end{pmatrix}. \quad (\text{C.9})$$

Se $\mathbf{a}^t$ é a primeira linha da matriz $\mathbf{A}$, podemos construir uma matriz

$$\mathbf{A}_1 = \mathbf{A} - \mathbf{x}_1 \mathbf{a}^t, \quad (\text{C.10})$$

onde $\mathbf{A}_1$ é uma matriz do tipo:

$$\mathbf{A}_1 = \begin{pmatrix} 0 & 0 & . & . & . & 0 \\ a'_{21} & a'_{22} & . & . & . & a'_{2n} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ a'_{n1} & a'_{n2} & . & . & . & a'_{nn} \end{pmatrix}. \quad (\text{C.11})$$

Seja $\mathbf{B}$ a matriz de ordem $n - 1$, extraída da matriz $\mathbf{A}_1$, eliminando-se sua primeira linha e também a primeira coluna. Assim:

$$\begin{pmatrix} a'_{22} & a'_{22} & . & . & . & a'_{2n} \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ a'_{n2} & a'_{n2} & . & . & . & a'_{nn} \end{pmatrix}. \quad (\text{C.12})$$

Agora pelo método das potências, calculamos o autovalor de maior valor absoluto da matriz $\mathbf{B}$. Seja λ_2 esse autovalor e $\mathbf{u}_2 = (u_{22}, \dots, u_{2n})^t$ seu autovetor correspondente.

Mostremos que:

- (i) λ_2 é autovalor de $\mathbf{A}_1$ e $\mathbf{z} = (0, u_{22}, \dots, u_{2n})^t$ seu autovetor correspondente.
- (ii) λ_2 é autovalor de $\mathbf{A}$ e $\mathbf{x}_2 = \mathbf{x}_1 + c\mathbf{z}$, onde $c = \frac{\lambda_2 - \lambda_1}{\mathbf{a}^t \mathbf{z}}$, seu autovetor correspondente.

Prova:

(i) Se λ_2 é autovalor de $\mathbf{B}$, então, λ_2 é raiz do polinômio característico de $\mathbf{B}$, portanto,

$|\mathbf{B} - \lambda_2 \mathbf{I}| = 0$. Mas,

$$|\mathbf{A}_1 - \lambda_2 \mathbf{I}| = \begin{vmatrix} -\lambda_2 & 0 & \dots & 0 \\ a'_{21} & a'_{22} - \lambda_2 & \dots & a'_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a'_{n1} & a'_{n2} & \dots & a'_{nn} - \lambda_2 \end{vmatrix} = -\lambda_2 |\mathbf{B} - \lambda_2 \mathbf{I}| = 0. \quad (\text{C.13})$$

Portanto, λ_2 é autovalor de $\mathbf{A}_1$. Ainda,

$$\begin{aligned} \mathbf{A}_1 \mathbf{z} &= \begin{pmatrix} 0 & 0 & \dots & 0 \\ a'_{21} & a'_{22} & \dots & a'_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a'_{n1} & a'_{n2} & \dots & a'_{nn} \end{pmatrix} \begin{pmatrix} 0 \\ u_{22} \\ \vdots \\ u_{2n} \end{pmatrix} = \begin{pmatrix} 0 \\ a'_{22}u_{22} + \dots + a'_{2n}u_{2n} \\ \vdots \\ a'_{n2}u_{22} + \dots + a'_{nn}u_{2n} \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ \lambda_2 u_{22} \\ \vdots \\ \lambda_2 u_{2n} \end{pmatrix} = \lambda_2 \begin{pmatrix} 0 \\ u_{22} \\ \vdots \\ u_{2n} \end{pmatrix} = \lambda_2 \mathbf{z}. \end{aligned} \quad (\text{C.14})$$

Portanto, $\mathbf{z} = (0, u_{22}, \dots, u_{2n})^t$ é autovetor de $\mathbf{A}_1$ associado ao autovalor λ_2 .

(ii) Mostremos que $\mathbf{A}\mathbf{x}_2 = \lambda_2 \mathbf{x}_2$, $\mathbf{x}_2 \neq 0$.

$$\text{como } \mathbf{x}_2 = \mathbf{x}_1 + c\mathbf{z} = \begin{pmatrix} 1 \\ u_{12} \\ \vdots \\ u_{1n} \end{pmatrix} + c \begin{pmatrix} 0 \\ u_{22} \\ \vdots \\ u_{2n} \end{pmatrix} = \begin{pmatrix} 1 \\ x_{22} \\ \vdots \\ x_{2n} \end{pmatrix} \quad (\text{C.15})$$

Temos a primeira componente de $\mathbf{u}_2 = 1$ logo $\mathbf{u}_2 \neq 0$.

Então:

$$\begin{aligned} \mathbf{A}\mathbf{x}_2 &= \mathbf{A}(\mathbf{x}_1 + c\mathbf{z}) = \mathbf{A}\mathbf{x}_1 + c\mathbf{z} = \mathbf{A}\mathbf{x}_1 + c\mathbf{A}\mathbf{z} = \lambda_1 \mathbf{x}_1 + c(\mathbf{A}_1 + \mathbf{x}_1 \mathbf{a}^t) \mathbf{z} \\ &= \lambda_1 \mathbf{x}_1 + c\mathbf{A}_1 \mathbf{z} + \frac{\lambda_2 - \lambda_1}{\mathbf{a}^t} \mathbf{x}_1 \mathbf{a}^t \mathbf{z} \\ &= \lambda_1 \mathbf{x}_1 + c\lambda_2 \mathbf{z} + \lambda_2 \mathbf{x}_1 - \lambda_1 \mathbf{x}_1 = \lambda_2 |\mathbf{x}_1 + c\mathbf{z}| = \lambda_2 \mathbf{x}_2. \end{aligned} \quad (\text{C.16})$$

Portanto, λ_2 é autovalor de $\mathbf{A}$ e $\mathbf{x}_2$ seu autovetor correspondente.

Apêndice D

Análise de estabilidade local de pontos fixos

Dado um sistema de equações diferenciais

$$\dot{\xi}_i = \Lambda_i(\xi, \mathbf{k}) \quad i = 1, 2, \dots, n. \quad (\text{D.1})$$

Devemos determinar todos os seus pontos de equilíbrio ou pontos fixos resolvendo o conjunto de equações algébricas resultante das condições $\dot{\xi} = 0$ para $i = 1, \dots, n$. Isto requer que encontremos todos os conjuntos de ξ_i tal que $\Lambda_i(\xi, \mathbf{k}) = 0$. Isto por si só, já é um problema complexo se as funções Λ_i são funções não-lineares. Se nenhuma solução exata pode ser encontrada, métodos numéricos devem ser empregados. Assumindo que podemos encontrar um ponto fixo, gostaríamos de determinar algo sobre sua natureza dinâmica. Ou seja, se encontramos um sistema natural próximo ou exatamente neste ponto, tenderia ele a permanecer próximo ao equilíbrio, ou se deslocaria para regiões distante deste ponto de equilíbrio? Em um sistema natural sempre haverá flutuações nos números da população, desta forma não podemos assumir que se o sistema é posto precisamente em um dos pontos fixos, ele automaticamente permanecerá neste ponto durante todo o tempo. Embora, não seja levada em consideração essas flutuações nas equações dinâmicas, devemos ter uma idéia de seus efeitos em sistemas que estão próximo ao equilíbrio por meio de *teoria de perturbação* ou *análise de estabilidade local*. Se as funções Λ_i que governam

a dinâmica do sistema são supostas contínuas, podemos usar uma expansão de Taylor de primeira ordem próxima ao ponto fixo. Desta forma, devemos substituir o sistema original de equações diferenciais não-lineares por um sistema de equações lineares que possui uma solução exata. Essas soluções nos fornecem, na maioria dos casos, uma indicação do comportamento da solução não-linear próximo ao ponto crítico.

Seja $\bar{\xi}_i$ o conjunto de valores em um ponto fixo, de forma que $\Lambda_i(\bar{\xi}_i) = 0$. Então, para os valores de ξ_i próximos a $\bar{\xi}_i$, a seguinte expressão se aplica:

$$\dot{\xi}_i = \Lambda_i(\bar{\xi}) + \sum_j \left. \frac{\partial \Lambda_i}{\partial \xi_j} \right|_{\xi=\bar{\xi}_i} (\xi_j - \bar{\xi}_j) + \mathcal{O}\left((\xi - \bar{\xi}_i)^2\right), \quad (\text{D.2})$$

onde o símbolo $\mathcal{O}\left((\xi - \bar{\xi}_i)^2\right)$ indica termos da ordem ou maior que $(\xi - \bar{\xi}_i)^2$. Se o ponto ξ_i é próximo o suficiente de $\bar{\xi}_i$, esperamos que estes termos sejam suficientemente pequenos em comparação com o termo linear, de forma que podemos desprezá-los.

Agora definimos a variável $z_i \equiv \xi_i - \bar{\xi}_i$ para $i = 1, \dots, n$. Então, desde que as coordenadas do ponto fixo são constantes, $\dot{z}_i = \dot{\xi}_i$. Também definimos a matriz $\mathbf{B}$, ou Jacobiano, com elementos

$$B_{ij} = \left. \frac{\partial \Lambda_i}{\partial \xi_j} \right|_{\xi=\bar{\xi}_i} = \left. \frac{\Lambda_i}{\partial z_j} \right|_{\mathbf{z}=0}. \quad (\text{D.3})$$

Como por definição de um ponto fixo, $\Lambda_i(\bar{\xi}) = 0$, a Equação (D.2) torna-se

$$\dot{z}_i = \sum_j B_{ij} z_j + \mathcal{O}(z^2). \quad (\text{D.4})$$

Se ignoramos termos de ordem superior, podemos escrever esta equação como

$$\dot{\mathbf{z}} = \mathbf{B}\mathbf{z}. \quad (\text{D.5})$$

Este é um sistema linear de equações diferenciais que pode ser resolvido pela mesma técnica utilizada para a solução do modelo de quase-espécies. Assumimos que $\mathbf{B}$ seja diagonalizável, tal que podemos escrever

$$B = U\Omega U^{-1}, \quad (\text{D.6})$$

onde Ω é a matriz diagonal cujos elementos $\Omega_{ii} \equiv \omega_i$ são os autovalores de B (que podem, em geral, ser números complexos), e a i -ésima coluna de U é o autovetor de B correspondente ao autovalor ω_i . Podemos escrever a Equação (D.5) na forma

$$\mathbf{U}^{-1}\dot{\mathbf{z}} = \Omega(\mathbf{U}^{-1}\mathbf{z}). \quad (\text{D.7})$$

Introduzindo $\zeta \equiv U^{-1}\mathbf{z}$, obtemos um sistema de equações diferenciais desacopladas

$$\dot{\zeta}_i = \omega_i \zeta_i. \quad (\text{D.8})$$

Cada uma dessas equações tem uma solução da forma

$$\zeta_i = A_i e^{\omega_i t}, \quad (\text{D.9})$$

onde A_i é uma constante determinada pelas condições iniciais. A solução para z_i é então encontrada da relação

$$z_i = \sum_j U_{ij} \zeta_j, \quad (\text{D.10})$$

de forma que z_i é uma combinação linear de termos exponenciais. O comportamento dos termos individuais z_i como função do tempo é assim determinado pelos autovalores da matriz B . Há três casos a considerar:

1. $\Re\omega_i < 0$ (onde $\Re$ denota a parte real de um número). Neste caso, o correspondente ζ_i decairá exponencialmente.
2. $\Re\omega_i > 0$. A magnitude do correspondente ζ_i crescerá exponencialmente.
3. $\Re\omega_i = 0$. Neste caso, Se $\Im\omega_i \neq 0$ ($\Im$ denota parte imaginária de um número), ζ_i oscilará com amplitude constante.

Vemos portanto que, se todos os autovalores satisfazem a condição $\Re\omega_i < 0$, então todos os ζ_i decairão a zero, e assim também todos os z_i . Neste caso, um pequeno

deslocamento do ponto fixo será atenuado no tempo e o sistema retornará ao ponto fixo. Tal ponto fixo é denominado de ponto fixo *assintoticamente estável*.

Se qualquer um dos autovalores satisfaz a condição $\Re \omega_i > 0$, o correspondente ζ_i crescerá exponencialmente. Qualquer elemento z_j que apresenta uma dependência em ζ_i também aumentará. Portanto, qualquer pequeno distúrbio do ponto fixo com componente na direção de z_j será amplificado, e neste caso o ponto fixo é dito *instável*.

É possível para um ponto fixo possuir uma mistura dessas duas propriedades. Se alguns dos autovalores têm partes reais negativas, e outros partes reais positivas, então é possível que alguns dos z_j contenham apenas termos com exponenciais negativos, enquanto outros têm no mínimo um termo com exponencial positiva. Assim, se uma perturbação ocorre de tal forma que apenas um z_j com termo exponencial negativo deve ser perturbado, o sistema retornará ao equilíbrio. Naturalmente, se qualquer um dos z_j com exponenciais positivas fosse perturbado, o distúrbio cresceria e o sistema continuaria a se deslocar do equilíbrio. Um ponto com essa dupla propriedade é denominado *ponto de sela*. Para tal propósito, pontos de sela são tidos quase como pontos totalmente instáveis, desde que é praticamente impossível que as flutuações aleatórias ocorram de forma seletiva, tal que só excitem aqueles valores de z_j contendo exponenciais negativas.

O terceiro caso referido acima, quando $\Re \omega_i = 0$, infelizmente nos dá muito pouca informação sobre o comportamento do ponto fixo associado. O comportamento oscilatório do sistema linear em torno do ponto fixo não é geralmente espelhado no correspondente sistema não-linear.

Bibliografia

- [1] M. Eigen. *Naturwissenschaften*, **58**, 465, 1971.
- [2] S. Kirkpatrick e B. Selman. *Science*, **264**, 1297, 1994.
- [3] J. Felsenstein. *Genetics*, **78**, 737, 1974.
- [4] M. Eigen e P. Schuster. *Naturwissenschaften*, **64**, 541, 1977.
- [5] M. Eigen e P. Schuster. *Naturwissenschaften*, **65**, 341, 1978.
- [6] M. Eigen; C. K. Biebricker; M. Gebinoga e W. C. Gardiner. *Biochemistry*, **30**, 1105, 1991.
- [7] J. A. Doudna e J. W. Szostak. *Nature*, **339**, 519, 1989.
- [8] M. Eigen e P. Schuster. *Naturwissenschaften*, **65**, 7, 1978.
- [9] J. Hofbauer e K. Sigmund. *Dynamical Systems and the Theory of Evolution*. Cambridge University Press, 1988.
- [10] J. Hofbauer; P. Schuster; K. Sigmund e R. Wolff. *SIAM (Soc. Ind. Appl. Math.) J. Appl. Math.*, **38**, 282, 1980.
- [11] P. F. Stadler e P. Schuster. *J. Math. Biol.*, **30**, 597, 1992.
- [12] A. García-Tejedor; F. Mórán e F. Montero. *J. Theor. Biol.*, **127**, 393, 1987.
- [13] J. C. Nuno; M. A. Andrade; F. Mórán e F. Montero. *Bull. Math. Biol.*, **55**, 385, 1993.
- [14] J. C. Nuno; M. A. Andrade e F. Montero. *Bull. Math. Biol.*, **55**, 417, 1993.
- [15] J. Maynard Smith. *Nature*, **280**, 445, 1979.
- [16] M. C. Boerlijst e P. Hogeweg. *Physica D*, **48**, 17, 1991.
- [17] M. R. Michod. *Am. Zool.*, **23**, 15, 1983.

- [18] R. Donato. *J. Phys. I*, **6**, 445, 1996.
- [19] Eörs Szathmáry. *Endocyt. C. Res.*, **3**, 113, 1986.
- [20] Eörs Szathmáry e László Demeter. *J. Theor. Biol.*, **128**, 463, 1987.
- [21] David Grey; Vivian Hutson e Eörs Szathmáry. *Proc. R. Soc. Lond. B*, **262**, 29, 1995.
- [22] D. S. Wilson. *The Evolution of Populations and Communities*. Benjamin-Cummings, Menlo Park, 1980.
- [23] K. E. Van Holde. Alan R. Liss, editores, *The Origin of Life and Evolution*, 1980.
- [24] C. R. Woese. Alan R. Liss, editores, *The Origin of Life and Evolution*, 1980.
- [25] R. M. Towe. *Prec. Res.*, **16**, 1, 1981.
- [26] S. A. Kauffman e S. Levine. *J. Theor. Biol.*, **128**, 11, 1987.
- [27] J. Swetina e P. Schuster. *Biophys. Chem.*, **16**, 329, 1982.
- [28] P. R. A. Campos e J. F. Fontanari. *Phys. Rev. E*, **58**, 2664, 1998.
- [29] G. Woodcock e P. G. Higgs. *J. Theor. Biol.*, **179**, 61, 1996.
- [30] M. Novak e P. Schuster. *J. Theor. Biol.*, **137**, 375, 1989.
- [31] C. K. Biebricher; M. Eigen e W. C. Gardiner. *Biochemistry*, **22**, 2544, 1983.
- [32] P. Schuster e J. Swetina. *Bull. Math. Biol.*, **50**, 635, 1988.
- [33] I. Leuthausser. *J. Chem. Phys.*, **84**, 1884, 1986.
- [34] P. Tarazona. *Phys. Rev. A*, **45**, 6038, 1992.
- [35] Stefano Galluccio. *Phys. Rev. E*, **56**, 4526, 1997.
- [36] M. Eigen; J. McCaskill e P. Schuster. *Adv. Chem. Phys.*, **75**, 149, 1989.
- [37] K. Binder. *J. Comput. Phys.*, **59**, 1, 1985.
- [38] D. Stauffer e A. Aharony. *Introduction to Percolation Theory*. Taylor & Francis, 1991.
- [39] P. R. A. Campos e J. F. Fontanari. *J. Phys. A*, **32**, L1, 1999.
- [40] D. E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, Massachusetts, 1989.

- [41] H. J. Muller. *Mutation Res.*, **1**, 2, 1964.
- [42] A. T. Bernardes. *Physica A*, **230**, 156, 1996.
- [43] L. Peliti. *cond-mat/9712027*, 1997.
- [44] J. F. Kimura e J. F. Crow. *An Introduction to Population Genetics Theory*. Harper & Row Publishers, 1970.
- [45] Tânia Tomé e Mário José de Oliveira. *Dinâmica Estocástica e Irreversibilidade*. EdUSP, 2001.
- [46] Gisela Engeln-Müllges e Frank Uhlig. *Numerical Algorithms with C*. Springer-Verlag, 1996.
- [47] D. Alves e J. F. Fontanari. *Phys. Rev. E*, **54**, 4048, 1996.
- [48] H. R. Grumm. *Analysis and Computation of Equilibria and Regions of Stability*, volume 8. II ASA Conf. Proc., Luxemburg, 1975.
- [49] G. Rowe. *Theoretical Models in Biology*. Clarendon Press-Oxford, 1994.
- [50] F. Mórán e F. Montero. *Computers & Chemistry*, **8**, 303, 1984.
- [51] A. GarcíaTejedor; A. R. Castaño; F. Mórán e F. Montero. *J. Mol. Evol.*, **26**, 294, 1987.
- [52] B. M. R. Stadler e P. F. Stadler. *Bull. Math. Biol.*, **53**, 469, 1991.
- [53] P. F. Stadler. *Biosystems*, **26**, 1, 1991.
- [54] J. C. Nuno e P. Tarazona. *Bull. Math. Biol.*, **56**, 875, 1994.
- [55] P. R. A. Campos; J. F. Fontanari e P. F. Stadler. *Physical Review E*, **61**, 2996, 2000.
- [56] D. Alves; P. R. A. Campos; A. T. C. Silva e J. F. Fontanari. *Phys. Rev. E*, **63**, 011911, 2001.
- [57] S. A. Boorman e P. R. Levitt. *The Genetics of Altruism*. Academic, New York, 1980.
- [58] S. Wright. *Genetics*, **16**, 97, 1931.
- [59] S. Wright. *Genetics*, **28**, 114, 1943.
- [60] I. Eshel. *Theor. Popul. Biol.*, **3**, 258, 1972.
- [61] K. Aoki. *Evolution*, **36**, 832, 1982.

- [62] A. T. C. Silva e J. F. Fontanari. *Eur. Phys. J. B*, **7**, 385, 1999.
- [63] E. Szathmáry. *J. Theor. Biol.*, **157**, 383, 1992.
- [64] E. Szathmáry. *J. Theor. Biol.*, **165**, 341, 1993.
- [65] E. Szathmáry e J. Maynard Smith. *J. Theor. Biol.*, **187**, 555, 1997.
- [66] Steven H. Strogatz. *The Evolution of Populations and Communities*. Addison-Wesley, 1994.
- [67] Daniel L. Hartl e Andrew G. Clark. *Principles of Population Genetics*. Sinauer Associates, 1988.
- [68] R. Donato; L. Peliti e M. Serva. *Theory Biosci.*, **116**, 309, 1997.
- [69] M. Kimura. *Nature*, **217**, 624, 1968.
- [70] M. Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, 1983.
- [71] C. M. Reyds; P. F. Stadler e P. Schuster. *Bull. Math. Biol.*, **59**, 339, 1997.
- [72] S. Gravilets e J. Gravner. *J. Theor. Biol.*, **184**, 51, 1997.
- [73] M. C. Wright e G. F. Joyce. *Science*, **276**, 614, 1997.
- [74] L. F. Landweber e I. D. Pokrovskaya. *Proc. Natl. Acad. Sci.*, **96**, 173, 1999.
- [75] E. H. Ekland e D. P. Bartel. *Nature*, **382**, 373, 1996.
- [76] S. F. Elena e R. E. Lenski. *Nature*, **390**, 395, 1997.
- [77] J. A. G. M. De Visser; R. E. Hoekstra e H. van den Enden. *Evolution*, **51**, 1499, 1997.
- [78] C. Adami. *Introduction to Artificial Life*. Springer, New York, 1998.
- [79] R. E. Lenski; C. Ofria; T. C. Collier e C. Adami. *Nature*, **400**, 661, 1999.
- [80] C. E. Shannon e W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, 1949.
- [81] T. M. Cover e J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

- [82] C. O. Wilke; J. Wang; C. Ofria; R. E. Lenski e C. Adami. *Aceito para publicação em Nature*, 2001.
- [83] B. Derrida. *Phys. Rev. Lett.*, **45**, 79, 1980.
- [84] B. Derrida. *Phys. Rep.*, **67**, 29, 1980.
- [85] B. Derrida. *Phys. Rev. B*, **24**, 2613, 1981.
- [86] S. A. Kauffman. *The Origins of Order, Self-Organization and Selection in Evolution*. Oxford University Press, 1993.
- [87] S. A. Kauffman. *Physica D*, **10**, 145, 1984.
- [88] S. Bonhoeffer e P. F. Stadler. *J. Theor. Biol.*, **164**, 359, 1993.
- [89] C. Reidys; C. V. Forst e P. Schuster. *Bull. Math. Biol.*, **63**, 57, 2001.
- [90] P. R. A. Campos; C. Wilke e C. Adami. *Em preparação*, 2001.
- [91] C. J. Thompson e J. L. McBride. *Math. Biosci.*, **21**, 127, 1974.
- [92] B. L. Jones; R. H. Enns e S. S. Rangnekar. *Bull. Math. Biol.*, **38**, 15, 1975.