

Universidade de São Paulo
Instituto de Física de São Carlos

**DINÂMICA DE MODELOS DE GENÉTICA
DE POPULAÇÕES COM RECOMBINAÇÃO**

Daniela Favarão Botelho

Dissertação apresentada ao Instituto de Física de São Carlos,
da Universidade de São Paulo, para a obtenção
do título de Mestre em Ciências: Física Básica.

Orientador: Prof. Dr. José Fernando Fontanari

✍

USP/IFSC/SBI



8-2-001510

São Carlos - 2003

IFSC - SBI
CLASS.....
CUTTER.....
TOMBO 70.271.....

Botelho, Daniela Favarão

Dinâmica de modelos de genética de populações e evolução/

Botelho, Daniela Favarão.

São Carlos, 2003. 70p.

Dissertação (mestrado) - Instituto de Física de São
Carlos, 2003.

Orientador: Prof. Dr. José Fernando Fontanari

1. Catraca de Muller. 2. Recombinação. 3. Balanço seleção-
mutação. I. Título.

**MEMBROS DA COMISSÃO JULGADORA DA DISSERTAÇÃO DE MESTRADO
DE DANIELA FAVARÃO BOTELHO, APRESENTADA AO INSTITUTO DE
FÍSICA DE SÃO CARLOS, DA UNIVERSIDADE DE SÃO PAULO, EM
20/03/2003.**

COMISSÃO JULGADORA:



Prof. Dr. José Fernando Fontanari (Orientador) – IFSC/USP



Prof. Dr. Domingos Alves – COCRP



Prof. Dr. Américo Tristão Bernardes – UFOP

Em memória de minha mãe, Romilde,
ao meu pai, Dorvalino, e aos meus irmãos,
Eduardo e Robson.
Como eu amo vocês!

**Este trabalho teve o apoio
financeiro da FAPESP.**

AGRADECIMENTOS

Agradeço a todas as pessoas que colaboraram direta ou indiretamente para que este trabalho fosse concluído:

- Ao professor Fontanari, pela orientação;
- A todos os colegas de grupo, em especial ao Léo e ao Colato - o que seria do meu mestrado sem vocês?
- Aos colegas de sala, Zé Paulo e Leonardo;
- Aos amigos, Miriam, Patricia, Norberto e Alexandre Yoshida - até perdi a conta das vezes que precisei de vossos ombros!
- Ao meu namorado, Nelson, pelo carinho e atenção e à sua família. Em especial a você, querido Urs...
- Mãe, Pai, Ró, Eduardo e Tia - obrigada por tudo. Vocês são o que a minha vida tem de melhor! Apesar de a Mãe não mais estar fisicamente presente entre nós, ela nos deixou muitas lembranças do amor, do carinho e da dedicação com que sempre nos cuidou, fazendo com que as nossas alegrias fossem também as suas. Nós nunca a esqueceremos!

Conteúdo

Lista de Figuras	viii
Lista de Tabelas	ix
1 Introdução	1
2 O Modelo	8
2.1 Descrição do modelo	9
3 A Catraca de Muller	16
4 O limite determinístico: população infinita	20
5 Simulações	26
6 Considerações Finais	50
Apêndices	52
A A distribuição de Poisson	53
B Solução analítica da dinâmica evolutiva	59
C Probabilidade de recombinantes	64

CONTEÚDO

Bibliografía

66

Lista de Figuras

2.1	Essência de um indivíduo da população que possui L genes com infinitos sítios. . .	10
2.2	Estas são as duas dinâmicas evolutivas estudadas. Na Dinâmica 1 são recombinados segmentos inteiros de indivíduos. Na Dinâmica 2, entretanto, os indivíduos são recombinados gene-a-gene. As frações localizadas nas figuras correspondem às probabilidades com que os indivíduos são escolhidos para transmitir seus genomas para a geração seguinte. A probabilidade de recombinação é dada por r e o acento circunflexo indica potência.	13
5.1	Dinâmica 1: número mínimo de mutações por indivíduo por geração mediados em 100 populações de $N = 100$ indivíduos para $s = 0.1$ (superiores) e $s = 0.01$ (inferiores), com $r = 1.0$. À esquerda, $U = \frac{0.5}{L}$, e à direita, $U = \frac{1.0}{L}$	28
5.2	Dinâmica 2: número mínimo de mutações por indivíduo por geração mediados em 100 populações de $N = 100$ indivíduos para $s = 0.1$ (superiores) e $s = 0.01$ (inferiores), com $r = 1.0$. À esquerda, $U = \frac{0.5}{L}$, e à direita, $U = \frac{1.0}{L}$	29
5.3	Dinâmica 1: R em função de L para $s = 0.01$ (à esquerda) e $s = 0.1$ (à direita). Na parte superior, $U = \frac{0.5}{L}$; na intermediária, $U = \frac{1.0}{L}$; e na inferior, $U = \frac{5.0}{L}$	35
5.4	Dinâmica 2: R em função de L para $s = 0.01$ (à esquerda) e $s = 0.1$ (à direita). Na parte superior, $U = \frac{0.5}{L}$; na intermediária, $U = \frac{1.0}{L}$; e na inferior, $U = \frac{5.0}{L}$	36

LISTA DE FIGURAS

- 5.5 Logaritmo da velocidade da catraca em função do tamanho N da população para $L = 1$ e $r = 1.0$. São variados os valores da taxa de mutação, U , e s é mantido fixo em 0.1. 38
- 5.6 Dinâmica 1: número mínimo de mutações por indivíduo em função das gerações mediados em 100 populações de $N = 100$ indivíduos, $L = 2$ e r variando. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$. À esquerda, $U = \frac{0.5}{L}$, e à direita, $U = \frac{1.0}{L}$ 39
- 5.7 Dinâmica 1: número mínimo de mutações por indivíduo em função das gerações mediados em 100 populações de $N = 100$ indivíduos, $L = 1000$ e r variando. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$. À esquerda, $U = \frac{0.5}{L}$, e à direita, $U = \frac{1.0}{L}$ 40
- 5.8 Dinâmica 2: número mínimo de mutações por indivíduo em função das gerações mediados em 100 populações de $N = 100$ indivíduos, $L = 2$ e r variando. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$. À esquerda, $U = \frac{0.5}{L}$, e à direita, $U = \frac{1.0}{L}$ 41
- 5.9 Dinâmica 2: número mínimo de mutações por indivíduo em função das gerações mediados em 100 populações de $N = 100$ indivíduos, $L = 1000$ e r variando. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$. À esquerda, $U = \frac{0.5}{L}$, e à direita, $U = \frac{1.0}{L}$ 42
- 5.10 Dinâmica 1: velocidade da catraca em função do fator de recombinação r para L variando de 2 a 100. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$, com $U = \frac{0.5}{L}$ à esquerda e $U = \frac{1.0}{L}$ à direita. Médias feitas em $NPOP = 50$ populações de $N = 100$ indivíduos. 44

LISTA DE FIGURAS

5.11 Dinâmica 2: velocidade da catraca em função do fator de recombinação r para L variando de 2 a 100. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$, com $U = \frac{0.5}{L}$ à esquerda e $U = \frac{1.0}{L}$ à direita. Médias feitas em $NPOP = 100$ populações de $N = 100$ indivíduos.	45
5.12 Dinâmica 1: velocidade da catraca em função do inverso de L para r variando de 0.1 a 1.0. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$, com $U = \frac{0.5}{L}$ à direita e $U = \frac{1.0}{L}$ à esquerda. Médias feitas em $NPOP = 100$ populações de $N = 100$ indivíduos.	47
5.13 Dinâmica 2: velocidade da catraca em função do inverso de L para r variando de 0.1 a 1.0. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$, com $U = \frac{0.5}{L}$ à esquerda e $U = \frac{1.0}{L}$ à direita. Médias feitas em $NPOP = 100$ populações de $N = 100$ indivíduos.	48
5.14 Dinâmica 2: Regressão linear da velocidade da catraca em função do inverso de L para r variando de 0.1 a 1.0. O parâmetro seletivo é $s = 0.1$ e $U = \frac{1.0}{L}$	49
A.1 Distribuição de Poisson com média 0.1 (à esquerda) e média 1 (à direita).	58
A.2 Distribuição de Poisson com média 10.	58

Lista de Tabelas

5.1	Velocidade da Catraca de Muller em função de L e U para $r = 1.0$ e $N = 10$ mediados em $NPOP = 100$ populações. Acima, Dinâmica 1. Abaixo, Dinâmica 2.	31
5.2	Velocidade da Catraca de Muller em função de L e U para $r = 1.0$ e $N = 50$ mediados em $NPOP = 100$ populações. Acima, Dinâmica 1. Abaixo, Dinâmica 2.	32
5.3	Velocidade da Catraca de Muller em função de L e U para $r = 1.0$ e $N = 100$ mediados em $NPOP = 100$ populações. Para os termos destacados com (*), $NPOP = 500$. Acima, Dinâmica 1. Abaixo, Dinâmica 2.	33
5.4	Velocidade da Catraca de Muller em função de L e U para $r = 1.0$ e $N = 200$ mediados em $NPOP = 100$ populações. Acima, Dinâmica 1. Abaixo, Dinâmica 2.	34
5.5	Velocidade da Catraca de Muller em função de N , U e s para $r = 1.0$ e $L = 2$ mediados em $NPOP = 10$ populações. Dados apenas da Dinâmica 1.	37

Resumo

Juntamente com o processo de mutação, a recombinação intragênica, vista como a troca recíproca de material genético entre genomas, é um dos principais fatores geradores da diversidade genética. De fato, os diversos mecanismos de recombinação existentes na natureza (sexo, por exemplo) são freqüentemente citados como invenções do processo de evolução via seleção natural para combater o efeito acumulativo de mutações deletérias, responsável pelo decréscimo gradual, mas contínuo, da adaptação (*fitness*) média de populações assexuadas de tamanho finito, num processo conhecido como catraca de Muller. Nesta dissertação, investigamos, através de simulações numéricas, como os mecanismos de recombinação afetam a velocidade da catraca de Muller em situações em que o efeito das mutações é multiplicativo, ou seja, o efeito deletério de uma nova mutação em um indivíduo independe das mutações anteriores. Trabalhamos com indivíduos haplóides de L genes que se reproduzem sexuadamente e estão sob a ação da seleção. Investigamos analiticamente o caso limite de população infinita e $L = 2$ genes, em que a catraca não atua. Para o caso específico de $L = 1$ onde, por construção, a recombinação não ocorre, derivamos a solução analítica da dinâmica evolutiva para qualquer tempo. De uma maneira geral, verificamos que a velocidade da catraca é retardada pelo acréscimo da taxa de recombinação. Em alguns dos casos estudados, essa velocidade tende a ser nula, indicando que o sexo pode evitar que populações sujeitas à ação de mutações deletérias sejam extintas. Nossos resultados numéricos também mostram que, como esperado, a movimentação da catraca tende a diminuir à medida que o tamanho da população aumenta.

Abstract

Together with the mutation process, the intragenic recombination, viewed as the reciprocal exchange of genetic material between genomes, is one of the main factors responsible for genetic diversity. Indeed, the mechanisms of recombination existing in nature (e.g., sex) are frequently cited as inventions of the evolution process via natural selection to avoid the cumulative effect of deleterious mutations, responsible for the gradual but continuous decline in mean fitness of finite asexual populations, in a process known as Muller's ratchet. In this dissertation, we investigate, through numerical simulations, how the recombination mechanisms affect the rate of Muller's ratchet in situations in which the effect of mutations is multiplicative, that is, the deleterious effect of a new mutation in an individual does not depend on the mutations it already carries. We work with haploid individuals of L genes which reproduce sexually and are under the effect of selection and recombination. We analytically investigate the limit case of infinite population and $L = 2$ genes, where the ratchet does not operate. For the specific case in which $L = 1$ where, by construction, recombination doesn't occur, we derive the analytical solution of the evolution dynamics for any time. In general, we verify that the ratchet's rate is retarded by the increase in the recombination rate. In some of the cases we studied, this rate tends to be null, indicating that sex may provide means to avoid extinction of populations subjected to the action of deleterious mutations. Our numerical results also show, as expected, that the ratchet's rate tends to slow down according to the increase in population size.

Capítulo 1

Introdução

A genética de populações trabalha com as leis e mecanismos da hereditariedade, bem como com as várias forças que possam alterar a composição genética de uma população. Ela tem sido essencial para a compreensão de diversos problemas no campo da biologia - como o estudo de processos evolutivos e transmissão de caracteres entre gerações. Em especial no aspecto matemático, a genética de populações vem elucidando diversas dúvidas e recriando conceitos. Seu desenvolvimento data da redescoberta do Mendelismo, em 1900, quando três botânicos, Karl Correns, Hugo De Vries e Eric von Tschermak-Seysenegg, a partir de experimentos envolvendo cruzamentos com plantas, chegaram independentemente às mesmas conclusões a que Mendel havia chegado em 1866, ou seja, depois de 34 anos [1]. Apenas a partir desse momento foi possível o estabelecimento de ligações adequadas entre a teoria da evolução, publicada por Charles Darwin em 1859, e a mecânica da hereditariedade.

Antes disso, um primo de Darwin, Francis Galton, dedicou-se ao estudo de caracteres quantitativos ¹ através da aplicação de métodos estatísticos para desvendar os princípios da hereditarie-

¹A variação observável de um caráter é geralmente quantitativa ou polimórfica. Quando existe uma série contínua ou gradual de tipos morfológicos - como os baseados em tamanho, peso e cor de pele em humanos - o modelo é quantitativo. Quando classes distintas de morfos existem dentro de uma população - como o dimorfismo sexual e tipos de grupos sanguíneos - o modelo é polimórfico [2].

1 Introdução

dade. Ele publicou, em 1897, sua teoria da ancestralidade, que viria a causar grande resistência ao mendelismo, a ser redescoberto em poucos anos. Para Galton, os genótipos dos indivíduos se misturavam a cada geração devido à ação do sexo. Assim, um pai de estatura alta e uma mãe de estatura baixa originariam um filho de estatura mediana (entre a paterna e a materna). Da mesma forma, um pai negro e uma mãe branca dariam origem a um filho de cor de pele entre branca e negra, isto é, mulata. A continuidade desse processo por milhares de anos resultaria na homogeneidade genética, na qual todos os indivíduos teriam genótipos idênticos. Em outras palavras, implicaria na perda da variedade. Mas não era isso que se observava: um casal de pai negro e mãe branca podiam tanto dar origem a um filho branco, negro, quanto mulato, de tal forma que a teoria de Galton não poderia valer para explicar a manutenção da variedade. Apesar disso, suas idéias de evolução logo foram difundidas, prejudicando a aceitação das idéias de Mendel sobre a hereditariedade, o que geraria uma disputa de natureza científica que duraria as primeiras décadas do século XX.

No começo do século passado, portanto, as descobertas no campo da genética iniciadas com a divulgação dos trabalhos de Mendel tiveram grande impacto sobre a teoria da evolução de Darwin, o qual acreditava que a evolução por seleção natural era um processo gradual que ocorria por meio da variação genética contínua². Os estudos dessa época mostraram que a base da evolução eram os genes e, igualmente importantes, seus mecanismos de transmissão de uma geração para outra. A partir dessa nova abordagem, novos conhecimentos e teorias foram sendo gerados, culminando no que chamamos hoje de “genética de populações”.

A genética de populações estuda as manifestações da herança a um nível populacional. Ela trabalha com modelos, ou seja, representações simplificadas de um mundo mais complexo: a realidade. Para isso, ela utiliza genes, gametas, etc, que são elementos da evolução sujeitos a regras operacionais previamente estabelecidas, com o intuito de se traduzir os fenômenos estudados. Es-

²Uma das implicações do trabalho de Mendel era que a evolução poderia ocorrer discretamente, isto é, não seria decorrente de uma variação contínua dos caracteres.

1 Introdução

As regras operacionais seguem, geralmente, princípios matemáticos e estatísticos, de modo que os modelos assim desenvolvidos são chamados modelos matemáticos. Os modelos têm a importância de estimar parâmetros e descrever fatos presentes, sempre voltados à previsões de comportamentos que podem ser testados experimentalmente. Eles permitem, portanto, um tratamento quantitativo dos fenômenos, o estabelecimento de previsões e também uma maneira de testar hipóteses.

Foi na revista *Science*, em 10 de julho de 1908, uma das primeiras publicações em genética de populações, por Godfrey Harold Hardy. O trabalho era resultante das questões levantadas numa conferência por um famoso estatístico da época, Yule, e havia sido apresentado a Hardy por intermédio do geneticista Punnett, o organizador da conferência. Mais tarde, verificou-se que o mesmo resultado de Hardy já havia sido publicado em 13 de janeiro do mesmo ano por um médico alemão, Wilhelm Weinberg e é conhecido na literatura atual como “princípio de Hardy-Weinberg”, que governa a distribuição de frequências genotípicas de uma população em equilíbrio na ausência de seleção e mutação na qual os cruzamentos são aleatórios, e tornou-se o ponto fundamental do desenvolvimento da genética de populações [3].

Um outro aspecto despertou o interesse dos geneticistas da época: o efeito do endocruzamento (cruzamento direcionado³) na distribuição das frequências genotípicas. O assunto foi tratado por H. S. Jennings e R. Pearl em uma série de trabalhos publicados entre 1912 e 1916, mas Sewall Wright, duvidando da veracidade de alguns dos resultados de Pearl, acabou desenvolvendo o problema, ampliando-o não somente para endocruzamentos como também para sistemas de cruzamento de um modo geral, o que culminou, em 1921, com diversas publicações intituladas “Systems of Mating”.

Juntamente com Ronald Alymer Fisher e John Burdon Sanderson Haldane, Wright tornou-se um dos mais importantes teóricos de genética de populações. Atualmente é bastante conhecido o

³É através dele que o homem promove uma seleção direcional escolhendo indivíduos portadores de certas características e promove o cruzamento entre eles. Desta forma, os genes responsáveis pelas características escolhidas têm aumentadas suas frequências e tendem a entrar em homozigose.

1 Introdução

modelo de evolução de Wright-Fisher, e foi por ele que iniciei os estudos do mestrado.

Assim, a essência do processo evolutivo é retratada pela genética de populações que, por volta da década de 30, já tinha suas bases completamente estabelecidas, quantificando duas propriedades importantes de populações biológicas: a capacidade de evoluir e de manter variação em longos períodos. publicou um artigo na revista *Genetics* intitulado “Evolution in Mendelian Populations” e o livro “The Causes of Evolution”.

No fim da década de 60 e início da de 70, a genética de populações sofreu mais um grande impacto. Pela primeira vez, a disponibilidade de dados experimentais do seqüenciamento de proteínas possibilitou a avaliação das teorias existentes no processo de substituição gênica. Com sua teoria neutralista da evolução molecular, Motoo Kimura [4] propôs que a evolução ocorria majoritariamente através da fixação de mutações neutras ou quase neutras, ou seja, daquelas que não trazem variação significativa na adaptação dos indivíduos da população.

Em 1964, Muller publicou um artigo em que descrevia o processo de perda de classes cromossômicas em populações finitas em que não ocorrem mutações reversas nem recombinação, considerando todas as mutações deletérias [5]. Seu trabalho teve grande importância em genética de populações pois abordava a acumulação de mutações em populações finitas, processo que ficou conhecido como *Catraca de Muller* e é o objeto principal de estudo desta dissertação ⁴.

A *Catraca de Muller* é um processo de perda progressiva dos indivíduos com menor número de mutações numa população, ou seja, dos mais adaptados. As mutações são consideradas alterações genéticas que, em geral, prejudicam a adaptação dos indivíduos portadores, diminuindo sua *fitness* (ou adaptabilidade), sobremaneira importante no processo seletivo. Em geral, se as mutações deletérias não forem detectadas por mecanismos de erro e reparo dos organismos vivos (talvez pela ausência deles em organismos mais rudimentares), elas tenderão a se acumular na população [8]. Esse acúmulo, entretanto, não pode ser ilimitado, pois as mutações muito prejudiciais tendem a ser removidas pelo processo de seleção [9]. Quando existe balanço entre mutação e seleção, a

⁴Para maiores detalhes da evolução da genética de populações, consultar [6, 7]

1 Introdução

distribuição de frequência do número de mutações por genoma atinge um estado de equilíbrio e a população não sofre mais deterioração [10]. É importante frisar, entretanto, que este argumento só vale para populações infinitas. Numa população finita, a Catraca começa a agir e o processo é essencialmente irreversível, já que a velocidade de atuação das mutações reversas não consegue reconstituir os genomas originais com nenhuma ou poucas mutações [8].

O processo de seleção compara as funções adaptação de cada um dos indivíduos de uma população numa dada geração e escolhe, com probabilidade proporcional à *fitness*, os indivíduos para cruzamento. Porque a *fitness* (ao menos no caso estudado por Muller) é função multiplicativa das mutações, quanto maior o número de mutações, menor a probabilidade de reprodução. O parâmetro que mede o grau deletério das mutações é chamado parâmetro seletivo e varia de 0 a 1, sendo 0 para dano mínimo e 1 para dano máximo.

Em populações assexuadas, a Catraca de Muller pode levar à extinção se permitimos que o tamanho da população varie. Para evitar que isso aconteça, a recombinação gênica, mistura dos genes do pai com os da mãe no filho, torna-se fator de elevada importância, pois pode criar na população indivíduos com pequeno número de mutações, que serão beneficiados no processo seletivo. Nesse aspecto, a influência da recombinação na Catraca deve ser considerada com atenção, pois pode reduzir - e até parar - a perda dos indivíduos mais adaptados, mantendo viável uma população que, sem este mecanismo, seria levada à extinção [11]. Fatores como migrações, catástrofes e expansões demográficas, entre outros, também devem ser considerados com atenção. Geralmente eles são ignorados em modelos teóricos - e é o que faremos - porque apesar de enriquecerem o estudo, trazem complexidade em demasia para o mesmo.

Foram observadas experimentalmente atuações da Catraca tanto em protozoários [11] como em vírus de RNA ($\phi 6$ e VSV - vesicular stomatitis virus) [12]. Os vírus de RNA são propícios a esse tipo de estudo por possuírem: altas taxas de mutação e fecundidade (não é preciso esperar muito tempo para que a população evolua e se reproduza); grandes tamanhos de população (o risco da perda de indivíduos pelo processo de amostragem é reduzido); e modos variados de reprodução,

1 Introdução

inclusive sexuada, em que há recombinação.

Nesse contexto geral, o objetivo deste trabalho é estudar as diversas formas de atuação da *Catraca*, mais especificamente a influência da recombinação no seu comportamento. Para isso, esta dissertação está dividida em 6 capítulos e 3 apêndices. No Capítulo 2, desenvolvemos o modelo a ser utilizado, justificando as simplificações necessárias decorrentes da complexidade descrita há pouco. É nele que definimos os mecanismos de hereditariedade, recombinação, seleção e reprodução, bem como a função adaptação, ingrediente fundamental para a evolução no modelo proposto.

O Capítulo 3 descreve a *Catraca* de Muller, que ocorre somente para populações finitas. Uma abordagem determinística é encontrada no Capítulo 4, onde são calculados os valores analíticos das concentrações de indivíduos com números fixos de mutações quando a população, já passado o transiente, atinge o equilíbrio. Neste capítulo, desenvolvemos um método de análise das concentrações para uma situação singular, na qual os indivíduos que compõem a população são haplóides e possuem dois genes, cada um com infinitos sítios.

As simulações do modelo descrito no Capítulo 2 se encontram no Capítulo 5. Alguns parâmetros são variados a fim de analisarmos suas influências na velocidade da *Catraca*. Entre eles, destacam-se: a taxa de recombinação, a taxa de mutação por indivíduo, o parâmetro seletivo (que mede o grau de dano das mutações), o tamanho do genótipo dos indivíduos e o tamanho da população. Nesse último caso, devido a limitações computacionais de tempo de execução do programa, o número máximo de indivíduos é de 5000.

O Apêndice A reserva uma descrição mais detalhada da maneira como um gerador de números aleatórios seguindo uma distribuição uniforme no intervalo de 0 a 1 é utilizado para a obtenção de números aleatórios que obedecem à distribuição de Poisson. No Apêndice B, encontra-se a dedução da solução da dinâmica evolutiva para uma população haplóide assexuada seguindo o modelo de Wright-Fisher com relevo multiplicativo. O problema foi abordado por um estudante de doutorado do grupo, Leonardo Paulo Maia, que fez uso das funções geratrizes e tornou elegante a obtenção

1 Introdução

do resultado lá exposto, que pode ser útil em diversas situações experimentais, em especial em populações que levam muito tempo para atingir o estado de equilíbrio, sendo observadas por pequenos intervalos de tempo apenas. O Apêndice C, por fim, destina-se à obtenção analítica das probabilidades de representação dos indivíduos na geração $t+1$ dada a representação da população na geração t .

Capítulo 2

O Modelo

A model is a mathematical abstraction which, although it cannot usually embody all the complexities of the real world, nevertheless aims to capture the essential features of reality in a sufficiently simplified form so that a mathematical analysis of the model is possible.

Warren Ewens [13]

Os modelos em genética de populações são fartos em simplificações. Como exemplo, mencionamos que as *fitness* são consideradas constantes, que as populações têm tamanho infinito ou que os acasalamentos são aleatórios. Modelos mais fiéis e próximos à realidade deveriam considerar, além desses fatores, a distribuição geográfica dos indivíduos, a mutação, migração, seleção e os variados sistemas de acasalamento, entre outros. O modo como esses fatores interagem, entretanto, é muito complexo e, apesar da busca em compreender suas implicações conjuntas numa população, eles não podem ser tratados simultaneamente, exigindo simplificações que abordem, a cada modelo, apenas um ou dois aspectos do mundo real.

Os modelos computacionais e matemáticos em genética de populações procuram estabelecer relações entre quantidades mensuráveis - os parâmetros - e associações matemáticas. Eles podem revelar os parâmetros relevantes de um sistema, testar hipóteses e servir de guia para a compre-

2.1 Descrição do modelo

ensão e análise de dados experimentais. Em especial, podem fazer previsões sobre o comportamento de um determinado sistema e testá-las através de observações. Neste capítulo, descrevemos o modelo utilizado tentando incluir todas as suas características essenciais e excluir as não essenciais, procurando manter um compromisso entre a facilidade de manipulação computacional e a realidade.

2.1 Descrição do modelo

A model is supposed to be simple: If it is not simpler than the real situation, then it is not a model.

Hartl and Clark [14]

Estudamos um modelo de evolução de populações compostas por indivíduos haplóides que obedecem regras probabilísticas de reprodução, mutação e recombinação. Por haplóides, entendemos indivíduos cujas características genéticas sejam dadas por apenas um cromossomo, sendo natural a não existência de dominância ou recessividade.

Os indivíduos que compõem a população são formados de L genes, cada um deles com infinitos sítios (ver Figura 2.1). Eles são descritos por sua carga genética contida nos L genes, que é transmitida de geração em geração através do processo de reprodução. O tamanho N da população é mantido constante, implicando numa capacidade limitada de indivíduos que podem ser suportados pelo ambiente e impondo uma competição efetiva entre eles. A variação na capacidade de adaptação dos indivíduos é um dos requisitos essenciais para que ocorra a evolução. Vamos supor, aqui, que essa capacidade dependa somente da seqüência que descreve o genótipo do indivíduo e as variações sejam resultado das mutações e recombinações que ocorrem durante o processo de reprodução.

A probabilidade de termos k novas mutações num indivíduo da população é dada por uma

2.1 Descrição do modelo

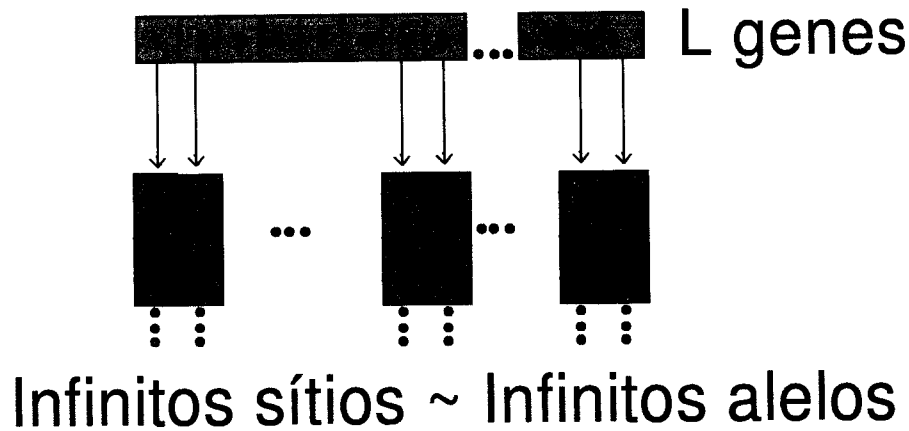


Figura 2.1: Essência de um indivíduo da população que possui L genes com infinitos sítios.

distribuição de Poisson,

$$M_k = e^{-U} \frac{U^k}{k!}, \quad (2.1)$$

em que U é o número médio de mutações por indivíduo [15]. De fato, no caso de tamanho de seqüência D e probabilidade de mutação por gene u , a probabilidade de k novas mutações segue uma distribuição binomial e quando $D \rightarrow \infty$ e $u \rightarrow 0$ o produto desses dois parâmetros, $Du = U$, fornece a taxa de mutação por seqüência e a distribuição binomial reduz-se à de Poisson¹.

Na natureza, as mutações que ocorrem nos indivíduos podem ser tanto benéficas quanto maléficas (ou deletérias). A proporção de mutações deletérias, no entanto, é muito maior do que as benéficas. É por este motivo que, em nosso modelo simplificado, vamos considerar deletérias todas as mutações, fazendo com que as chances de reprodução dos indivíduos sejam proporcionais ao número de mutações que eles possuam. A função que relaciona o número de mutações em um indivíduo com sua adaptação é a seguinte:

$$w_{kl\dots m} = (1 - s)^{k+l+\dots+m}, \quad (2.2)$$

em que os L índices k, l, \dots, m correspondem às mutações para cada um dos L genes do indivíduo

¹Para dedução, consultar Apêndice A.

2.1 Descrição do modelo

e $s \in [0, 1]$ é chamado parâmetro seletivo, medindo o grau de dano das $k + l + \dots + m$ mutações.

Com essa escolha temos

$$w_{kl\dots m} = (1 - s)^k (1 - s)^l \dots (1 - s)^m = w_k w_l \dots w_m \quad (2.3)$$

e o efeito total das mutações é simplesmente o produto do efeito de cada uma. Por essa razão, esse caso é conhecido como relevo de replicação multiplicativo, uma referência ao nome em inglês *multiplicative fitness landscape*.

A grandeza relevante para a determinação do destino de um indivíduo i na população é sua adaptação relativa W_i , definida como a razão entre a sua adaptação, $w_{(kl\dots m)_i}$, e a adaptação total da população na geração t , T_t ,

$$W_i = \frac{w_{(kl\dots m)_i}}{T_t}, \quad (2.4)$$

com

$$T_t = \sum_{j=1}^N (1 - s)^{(k+l+\dots+m)_j}, \quad (2.5)$$

onde k_j, l_j, \dots, m_j representam o número de mutações do 1º gene, 2º gene, ..., e L -ésimo gene, respectivamente, do indivíduo $j = 1, 2, 3, \dots, N$ da população. É importante observar que W_i depende da geração (tempo²) pois T_t depende da composição genética da população, que evolui no tempo.

À medida que o coeficiente seletivo s se aproxima de 1 ($s \rightarrow 1$) o efeito deletério das mutações torna-se cada vez maior, enquanto que para s pequeno as mutações praticamente não afetam a função adaptação. O caso em que $s = 0$ é denominado limite de seleção neutra, no qual as mutações não produzem efeito algum na capacidade reprodutiva dos indivíduos, de modo que todos têm probabilidades idênticas de se reproduzir. O limite $s = 1$ é chamado de regime de seleção extrema, no qual somente os indivíduos com nenhuma mutação conseguem se reproduzir e os demais surgem apenas através das mutações dessa classe de indivíduos [10].

²Usamos tempo e geração indistintamente neste trabalho.

2.1 Descrição do modelo

Definidas a função adaptação dos indivíduos e a probabilidade de mutação, é ainda necessário deprever os mecanismos de recombinação, também responsáveis por gerar diversidade na passagem de gerações. Para isso, vamos trabalhar agora com dois exemplos ilustrativos, a saber: Dinâmica 1 e Dinâmica 2, cujos esquemas podem ser visualizados na Figura 2.2 e são descritos a seguir.

- Dinâmica 1: Um casal com $L = 4$ genes (cada gene com um sítio que pode assumir apenas dois valores, alelos A e B), digamos $AAAA$ e $BBBB$, é escolhido da população no tempo t para produzir um filho que pertencerá à geração $t + 1$. Se o terceiro gene, escolhido aleatoriamente, é o de recombinação, são quatro os possíveis filhos deste casal: $AAAA$, $AAAB$, $BBBA$ e $BBBB$, dos quais somente um é passado para a geração seguinte. Definindo $r \in [0, 1]$ como uma probabilidade de recombinação e a um número aleatório entre 0 e 1, escolhemos indivíduos recombinantes se $a < r$, e indivíduos não-recombinantes se $a > r$. É evidente que dentre estes dois intervalos ainda surgem duas possibilidades de escolha, já que se $a < r$ podemos ter $AAAB$ ou $BBBA$. Este problema é facilmente solucionado quando dividimos o intervalo $[0, r]$ em dois: $[0, \frac{r}{2}]$ e $[\frac{r}{2}, r]$. Se $a < \frac{r}{2}$ escolhemos um dos dois recombinantes, caso contrário, o outro.

A determinação do gene de recombinação é aleatória e segue uma distribuição uniforme de probabilidade no intervalo de 1 a $L - 1$. O procedimento para a determinação deste gene, bem como para a recombinação, vale para qualquer L e é repetido até que o número de indivíduos na geração $t + 1$ provenientes deste mecanismo se iguale ao número de indivíduos da população parental. Na parte superior da Figura 2.2, está ilustrada a Dinâmica 1. As frações nela encontradas se referem às probabilidades de ocorrência de cada uma das seqüências. Esse procedimento para a recombinação é o empregado na literatura de algoritmo genético [16].

- Dinâmica 2: O mesmo casal da Dinâmica 1 é escolhido para gerar um filho para a geração seguinte. O processo de recombinação não mais acontece escolhendo-se um gene e mesclando-

2.1 Descrição do modelo

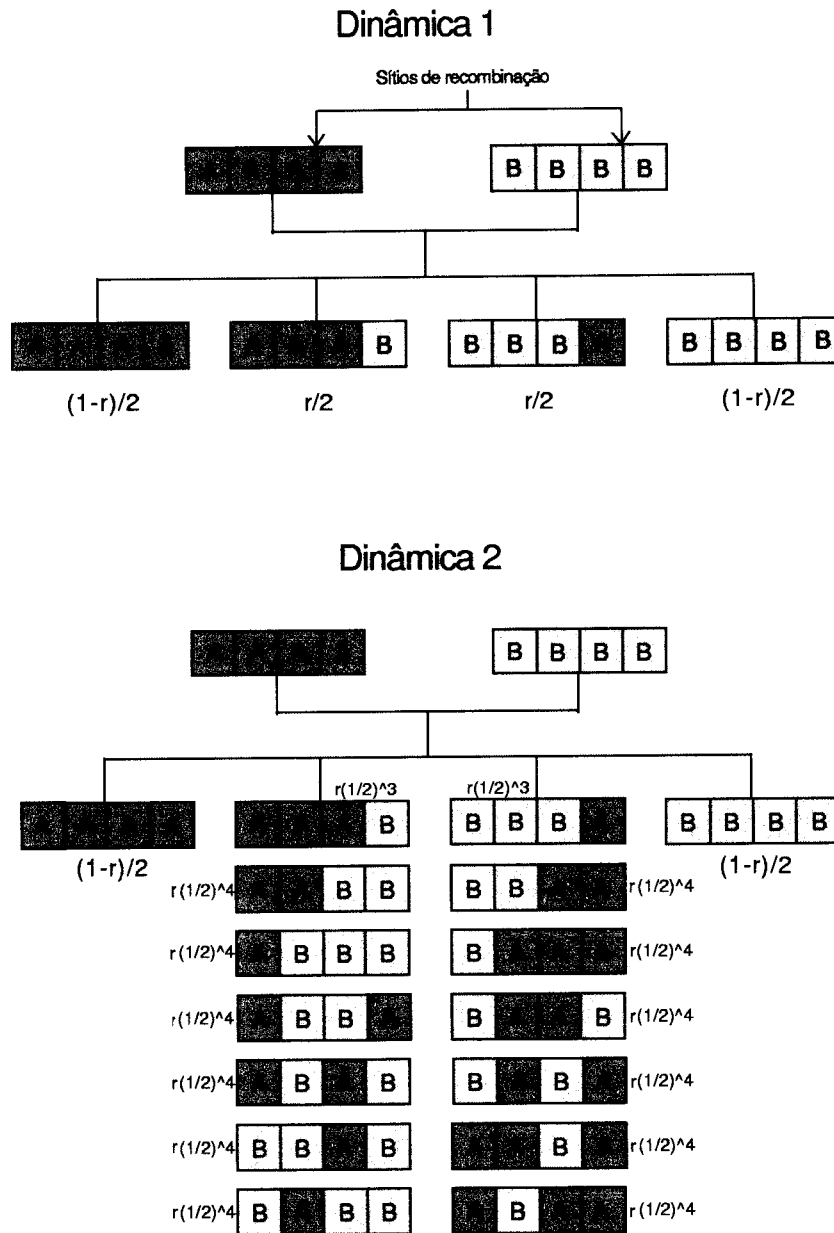


Figura 2.2: Estas são as duas dinâmicas evolutivas estudadas. Na Dinâmica 1 são recombinados segmentos inteiros de indivíduos. Na Dinâmica 2, entretanto, os indivíduos são recombinados gene-a-gene. As frações localizadas nas figuras correspondem às probabilidades com que os indivíduos são escolhidos para transmitir seus genomas para a geração seguinte. A probabilidade de recombinação é dada por r e o acento circunflexo indica potência.

2.1 Descrição do modelo

se as seqüências parciais dos genomas parentais. Apesar da probabilidade de recombinação continuar sendo r e haver somente duas possibilidades de filhos não-recombinantes, $AAAA$ e $BBBB$, cada um com probabilidade $\frac{1-r}{2}$, temos, agora, que dado que houve recombinação, isto é, $a < r$, cada gene pode vir tanto do pai quanto da mãe (com igual chance), incrementando o número de filhos diferentes que podem ser gerados. Deve-se frisar, no entanto, que entre os recombinantes, existem dois deles que são mais prováveis, $AAAB$ e $BBBA$, como pode ser verificado na parte inferior da Figura 2.2. O motivo dessa ocorrência é oriundo da dinâmica de programação utilizada, que escolhe, gene a gene, a sua procedência (se materna ou paterna) e beneficia as seqüências que são homogêneas (ou seja, só com A's ou só com B's) até o gene $L - 1$. Nesses dois casos, dado que houve a recombinação, o último gene deve ser de origem distinta à dos $L - 1$ genes. Todos os outros recombinantes, por sua vez, têm probabilidades idênticas de produção, como pode ser verificado através das frações na Figura 2.2. Uma formulação matemática mais detalhada para esse problema é apresentada no Apêndice C. Esse procedimento, em contraste com o utilizado em algoritmos genéticos, é utilizado na literatura de genética de populações [17, 18].

Estando definidas as dinâmicas parciais para a evolução, vamos, agora, descrever o processo como um todo. As gerações são discretas e a população na geração $t + 1$ é obtida da população na geração t através de N sorteios de pares de indivíduos, sendo eles escolhidos de acordo com a função adaptação relativa que, por sua vez, é função do número de mutações dos indivíduos. Esses N pares dão origem a N novos elementos provenientes de casamentos com - ou sem - recombinação, conforme explicado anteriormente nesta seção. Para cada um dos N indivíduos da nova geração e para cada um de seus L genes é escolhido o número de novas mutações através da distribuição de Poisson. Assim sendo, o número de mutações de um indivíduo é igual à soma do número de mutações herdadas de seus pais com as novas mutações geradas pela equação (2.1) para todos os L genes. Esse procedimento de obter as populações a cada geração por amostragem e reposição

2.1 Descrição do modelo

é conhecido na literatura de genética de populações como modelo de Wright-Fisher [19], sendo também bastante utilizado na literatura de algoritmos genéticos [14]. Deve-se notar que a cada geração toda a população é renovada, pois os pais dos indivíduos não convivem com os filhos, não havendo, portanto, superposição de gerações.

Uma das limitações do modelo proposto é considerar que todas as mutações são deletérias (isto é, diminuem a adaptação dos indivíduos). Apesar de ser uma suposição razoável [12, 20], um modelo mais realista deveria considerar seu efeito (deletério ou benéfico) de uma forma estocástica. Os efeitos das mutações são relevantes para o estudo do processo chamado *Catraca de Muller* [5], que consiste na perda estocástica sucessiva e irreparável dos indivíduos com menor número de mutações, ou seja, dos mais adaptados da população. Em populações assexuadas, a *Catraca de Muller* pode levar à degradação e eventual extinção da população. De fato, uma das teorias mais populares sobre a evolução do sexo baseia-se no fato de que em uma população sexuada não ocorre acumulação de mutações deletérias e, portanto, a *catraca* deixa de atuar [21]. É este o tópico que vamos abordar no próximo capítulo.

Capítulo 3

A Catraca de Muller

A Catraca de Muller - assim denominada em reconhecimento a H. J. Muller, que descreveu o fenômeno em seu artigo de 1964 [5] - é um processo de acumulações deletérias que pode levar à extinção de populações assexuadas através da diminuição contínua da adaptação média da população [22] .

Em uma população finita, a classe de indivíduos com menor número de mutações pode não ser representada na geração seguinte devido à ação da deriva genética, que são flutuações nas frequências de classes de indivíduos. Sem recombinação ou mutações reversas, as mutações deletérias se acumulam nos indivíduos, tornando-os cada vez menos adaptados. Nessas circunstâncias, a adaptação média da população entra em declínio e ela pode até ser extinta.

A catraca funciona da seguinte forma: seja N o número total de indivíduos na população e N_k , N_l, \dots, N_m o número de indivíduos com k, l, \dots, m mutações, sendo agora k, l, \dots, m os resultados das somas das mutações em todos os L genes de um indivíduo. Como em nosso modelo todas as mutações são deletérias, a adaptação dos indivíduos diminui com o acréscimo de k, l, \dots, m . O número de indivíduos na classe mais adaptada (com nenhuma mutação) é N_0 . Isto significa que nos L genes dos N_0 indivíduos não ocorreu nenhuma mutação. N_0 é uma variável aleatória que, devido às flutuações, eventualmente atinge o valor zero. Como, por hipótese do modelo, não

3 A Catraca de Muller

existem mutações reversas, a classe $k = 0$ é extinta da população e a classe mais adaptada passa a ser $k = 1$, com número de indivíduos $N_1 = N$. Essa classe é obtida quando a soma das mutações nos L genes de um indivíduo resulta num total de 1. Mais uma vez, o destino desses indivíduos é a extinção e a classe mais adaptada passa a ser aquela em que $k = 2$, e assim por diante. É evidente que seguindo este mecanismo as distribuições de N_k, N_l, \dots, N_m tendem a se mover gradativamente para maiores valores de k, l, \dots, m , similarmente ao movimento de uma catraca. A velocidade com que a catraca se move é medida através do coeficiente angular da curva de $k_{min} \times t$, sendo k_{min} o valor mínimo das somas das mutações de um indivíduo, e t medido em gerações. Assim, quanto maior a velocidade da catraca, maior será o número mínimo de mutações (k_{min}) acumuladas na população em um certo número de gerações.

Mutações extremamente prejudiciais são rapidamente eliminadas da população, pois são altamente perceptíveis através da função adaptação, que diminui drasticamente. Neste caso, como foi observado por Lynch e Gabriel, e Pamilo e colaboradores [24, 26], a acumulação de mutações é muito lenta. Mutações de efeito moderado, entretanto, são menos afetadas pelo efeito purificador da seleção e são transmitidas mais facilmente de geração para geração. O resultado desse processo é a rápida diminuição da adaptação média da população e o conseqüente acréscimo na velocidade da catraca.

Mutação, seleção e flutuação das classes de indivíduos são, portanto, fatores que movem e orientam a catraca. Populações pequenas, mas com altas taxas de mutação resultam numa grande flutuação de freqüências de N_k, N_l, \dots, N_m e acumulam mutações deletérias rapidamente. Já para populações grandes, as flutuações ocorrem com menor intensidade, resultando numa acumulação mais lenta de mutações deletérias. Ainda para populações grandes ($N \geq 1000$), maiores taxas de mutação facilitam a perda dos indivíduos mais aptos.

A velocidade com que a catraca avança também é sensível ao tamanho da população, pois está subordinada às flutuações das classes de indivíduos, fenômeno conhecido como deriva genética [12]. Em longos períodos, a acumulação de mutações deletérias apresenta sérios problemas para

3 A Catraca de Muller

qualquer organismo assexuado, exceto para aqueles com tamanhos grandes de população, como as bactérias [8]. Em nosso trabalho, o tamanho da população é mantido constante, embora esta seja "uma visão inconsistente com a suposição de que todas as mutações são prejudiciais" [27]. Ainda assim, poucos foram os autores que estudaram o efeito do acúmulo de mutações no tamanho da população [8, 11, 24, 25].

O processo de recombinação é outro fator que leva à maior diversidade genética e conseqüente criação de indivíduos mais, ou menos aptos na população. Por ser um processo estocástico, a união de gametas pode dar origem a indivíduos antes inexistentes na população, cujas adaptações relativas podem ser melhores ou piores daquelas já encontradas. As melhores seriam resultado da união de gametas com poucas mutações e as piores, de gametas com número substancial de mutações. Onde se encontra, então, a vantagem de haver recombinação no processo evolutivo? A resposta vem no passo seguinte à criação de novos indivíduos: na seleção. É evidente que indivíduos pouco adaptados terão poucas chances de passar para a geração posterior, sendo eliminados da população. Em contrapartida, os indivíduos mais bem adaptados terão chances maiores de serem escolhidos para a nova geração. Nota-se, portanto, que apesar da recombinação parecer momentaneamente prejudicial, ela é benéfica de modo geral, pois atua na melhoria da adaptação média da população, eliminando indivíduos com mutações deletérias em excesso. Em decorrência disso, a velocidade da catraca diminui.

Podemos, ainda, destacar dois fatores que alteram a velocidade da catraca: a taxa de recombinação e o número L de genes dos indivíduos da população. À medida que a taxa de recombinação aumenta fica mais fácil, probabilisticamente, de acontecerem acasalamentos com recombinação. Conseqüentemente, a velocidade da catraca diminui. A menos que a recombinação seja muito restrita ou que a taxa de mutação por gene seja muito alta, a presença de recombinação sempre retarda o progresso da catraca [22]. Esse comportamento também é observado se aumentamos o número L de genes de um indivíduo, pois estando as mutações mais distribuídas, torna-se mais provável que a união de gametas com poucas mutações ocorra. No próximo capítulo estudaremos

3 A Catraca de Muller

o caso limite de população infinita com $L = 2$, em que a catraca não mais atua.

Capítulo 4

O limite determinístico: população infinita

Considerando o modelo descrito no Capítulo 2, vamos, agora, investigar a situação em que os indivíduos são compostos de apenas dois genes, $L = 2$, e o tamanho da população é infinito. Nesse limite, a mutação e a recombinação, que juntas geram novos indivíduos, e a seleção, que elimina os menos aptos, balanceiam-se de modo que a dinâmica evolutiva atinge um estado estacionário, no qual as concentrações de indivíduos com k mutações totais não dependem do tempo. O objetivo deste capítulo é determinar os valores das concentrações no equilíbrio em termos do parâmetro seletivo s , da taxa de mutação por gene¹, $\frac{U}{L} = u$, e da probabilidade de recombinação r .

Vamos, inicialmente, separar os indivíduos em infinitas classes de j e k , j e k variando de zero a infinito, de acordo com o número de mutações que os indivíduos possuam no primeiro e segundo genes, respectivamente. Para isso, definamos $C_{jk}(t) = N_{jk}/N$, que representa a concentração de indivíduos com j e k mutações nos seus dois genes, cuja soma deve ser a unidade,

$$\sum_{jk} C_{jk}(t) = 1. \quad (4.1)$$

¹A taxa de mutação por indivíduo(formado por L genes) é U e, se $L = 1$, $U = u$.

4 O limite determinístico: população infinita

Nosso problema consiste em determinar $C_{jk}(t+1)$ dadas as concentrações $C_{jk}(t)$. Lembremos que os indivíduos da geração $t+1$ são escolhidos em N sorteios com reposição de casais retirados da população no tempo t .

Seja $P[mn, op]$ a probabilidade que um casal do tipo mn , op contribua na formação de um indivíduo na próxima geração. Temos

$$P[mn, op] = P[mn]P[op], \quad (4.2)$$

já que as retiradas são com reposição e a probabilidade de se escolher o indivíduo mn no primeiro sorteio e o indivíduo op no segundo sorteio são independentes. Deste modo,

$$P[mn, op] = \frac{C_{mn}(t)w_{mn}}{W_t} \frac{C_{op}(t)w_{op}}{W_t}, \quad (4.3)$$

onde w_{mn} e w_{op} são as adaptações absolutas dos indivíduos com genótipo mn e op , calculadas a partir da equação (2.3). Aqui,

$$W_t = \sum_{ij} C_{ij}(t)w_{ij} \quad (4.4)$$

é a adaptação média da população no tempo t .

Sendo escolhido um casal desse tipo, definimos $P[j'k'|mn, op]$ a probabilidade que um indivíduo $j'k'$ seja formado do acasalamento dos indivíduos mn e op . Como ainda não consideramos a mutação, a única maneira deste evento ocorrer, se não houver recombinação, é que pelo menos um dos pais seja de genótipo $j'k'$. Já no caso em que existe recombinação, é necessário que as seqüências j' e k' apareçam nos pais. Em formulação matemática, dado que r é a probabilidade de recombinação, temos

$$P[j'k'|mn, op] = \frac{1-r}{2}\delta_{j'm}\delta_{k'n} + \frac{1-r}{2}\delta_{j'o}\delta_{k'p} + \frac{r}{2}\delta_{j'm}\delta_{k'p} + \frac{r}{2}\delta_{j'o}\delta_{k'n}, \quad (4.5)$$

em que δ_{jm} é a delta de Kronecker.

Formado um indivíduo $j'k'$, é ainda preciso que j' mute para j e k' mute para k para que finalmente se obtenha um indivíduo jk . Nota-se, portanto, que j' e k' devem ser menores ou,

4 O limite determinístico: população infinita

no máximo, iguais a j e k , respectivamente. Como mencionado anteriormente, se o tamanho da seqüência é infinito e a taxa de mutação é pequena, essa probabilidade de mutação segue uma distribuição de Poisson,

$$P[jk|j'k'] = \frac{e^{-u}u^{j-j'} e^{-u}u^{k-k'}}{(j-j')! (k-k')!}, \quad (4.6)$$

onde novamente as probabilidades são multiplicadas por se tratarem de eventos independentes.

Assim, a probabilidade de encontrarmos uma concentração $C_{jk}(t+1)$ dados $C_{jk}(t)$ pode ser escrita como

$$C_{jk}(t+1) = \sum_{j'k'} \sum_{mn} \sum_{op} P[jk|j'k'] P[j'k'|mn, op] P[mn, op],$$

sendo as somas sobre todos os valores possíveis de j' , k' , m , n , o e p , respeitadas algumas restrições.

Temos

$$\begin{aligned} C_{jk}(t+1) &= \sum_{j'k'} \sum_{mn} \sum_{op} P[jk|j'k'] \left[\frac{1-r}{2} \delta_{j'k',mn} + \frac{1-r}{2} \delta_{j'k',op} \right. \\ &\quad \left. + \frac{r}{2} \delta_{j'k',mp} + \frac{r}{2} \delta_{j'k',on} \right] P[mn] P[op] \end{aligned} \quad (4.7)$$

e eliminando as deltas ao proceder com as somas em j' e k' , obtemos

$$\begin{aligned} C_{jk}(t+1) &= \sum_{mn} \sum_{op} \frac{1-r}{2} P[jk|mn] P[mn] P[op] \\ &\quad + \sum_{mn} \sum_{op} \frac{1-r}{2} P[jk|op] P[mn] P[op] \\ &\quad + \sum_{mn} \sum_{op} \frac{r}{2} P[jk|mp] P[mn] P[op] \\ &\quad + \sum_{mn} \sum_{op} \frac{r}{2} P[jk|on] P[mn] P[op]. \end{aligned} \quad (4.8)$$

Explicitando as distribuições de probabilidade encontramos

$$\begin{aligned} C_{jk}(t+1) &= (1-r) \sum_{mn} \frac{e^{-u}u^{j-m} e^{-u}u^{k-n}}{(j-m)! (k-n)!} \frac{C_{mn}(t)w_{mn}}{W_t} \\ &\quad + r \sum_{mn} \sum_{op} \frac{e^{-u}u^{j-m} e^{-u}u^{k-p}}{(j-m)! (k-p)!} \frac{C_{mn}(t)w_{mn}}{W_t} \frac{C_{op}(t)w_{op}}{W_t}. \end{aligned} \quad (4.9)$$

4 O limite determinístico: população infinita

Ou ainda,

$$C_{jk}(t+1) = (1-r) \sum_{mn} \frac{e^{-u} u^{j-m} e^{-u} u^{k-n} C_{mn}(t) w_{mn}}{(j-m)! (k-n)! W_t} + r \left(\sum_{mn} \frac{e^{-u} u^{j-m} C_{mn}(t) w_{mn}}{(j-m)! W_t} \right) \left(\sum_{mn} \frac{e^{-u} u^{k-n} C_{mn}(t) w_{mn}}{(k-n)! W_t} \right), \quad (4.10)$$

já que os índices são mudos e trocamos *op* por *mn* no último termo. Finalizando,

$$C_{jk}(t+1) = (1-r) \sum_{m \leq j} \sum_{n \leq k} P_{j-m} P_{k-n} \frac{C_{mn}(t) w_{mn}}{W_t} + r \left(\sum_{m \leq j} \sum_{n(\text{livre})} P_{j-m} \frac{C_{mn}(t) w_{mn}}{W_t} \right) \left(\sum_{m(\text{livre})} \sum_{n \leq k} P_{k-n} \frac{C_{mn}(t) w_{mn}}{W_t} \right), \quad (4.11)$$

com

$$P_{i-j} = \frac{e^{-u} u^{i-j}}{(i-j)!}. \quad (4.12)$$

Uma solução de (4.11) no equilíbrio é obtida quando supomos $C_{jk} = C_j C_k$, sendo C_j e C_k as concentrações estacionárias das mutações no primeiro e no segundo genes, respectivamente, pois assim, eliminando as dependências em t , temos

$$C_{jk} = (1-r) \sum_{m \leq j} \sum_{n \leq k} P_{j-m} P_{k-n} \frac{C_{mn} w_{mn}}{W} + r \left(\sum_{m \leq j} P_{j-m} \frac{C_m w_m}{\sum_m C_m w_m} \right) \left(\sum_{n \leq k} P_{k-n} \frac{C_n w_n}{\sum_n C_n w_n} \right), \quad (4.13)$$

onde usamos

$$W = \sum_{mn} C_{mn} w_{mn} = \sum_m C_m w_m \sum_n C_n w_n. \quad (4.14)$$

Ou seja,

$$C_{jk} = (1-r) \sum_{m \leq j} \sum_{n \leq k} P_{j-m} P_{k-n} \frac{C_{mn} w_{mn}}{W} + r \sum_{m \leq j} \sum_{n \leq k} P_{j-m} P_{k-n} \frac{C_{mn} w_{mn}}{W}. \quad (4.15)$$

Cancelando os termos em r ,

$$C_{jk} = \sum_{m \leq j} \sum_{n \leq k} P_{j-m} P_{k-n} \frac{C_{mn} w_{mn}}{W}. \quad (4.16)$$

4 O limite determinístico: população infinita

Notamos, portanto, que para a *fitness* escolhida, a solução de equilíbrio não depende do fator de recombinação r . Sendo assim, identificamos C_j e C_k

$$C_{jk} = \left(\sum_{m \leq j} P_{j-m} \frac{C_m w_m}{\sum_m C_m w_m} \right) \left(\sum_{n \leq k} P_{k-n} \frac{C_n w_n}{\sum_n C_n w_n} \right) = C_j C_k. \quad (4.17)$$

Se a *fitness* é $(1-s)^m$ para m mutações,

$$C_j = \sum_{m \leq j} P_{j-m} \frac{C_m (1-s)^m}{T_e}, \quad (4.18)$$

sendo $T_e = \sum_m C_m (1-s)^m$. Fazendo $j = 0$, obtemos

$$T_e = P_0 w_0 = P_0 = e^{-u}, \quad (4.19)$$

levando ao resultado de que T_e é independente de s , desde que C_0 não seja nulo.

Podemos obter uma solução explícita para C_j em termos das frequências das classes de maior adaptação,

$$C_j = \frac{1}{1 - (1-s)^j} \sum_{m=0}^{j-1} C_m (1-s)^m \frac{u^{j-m}}{(j-m)!}, \quad j = 1, 2, \dots \quad (4.20)$$

Substituindo seqüencialmente $j = 1, 2, \dots$ determinamos que

$$C_j = C_0 \frac{\theta^j}{j!}, \quad (4.21)$$

onde $\theta = \frac{u}{s}$ e $C_0 \neq 0$. Usando a condição de normalização $\sum_j C_j = 1$, obtemos $C_0 = e^{-\theta}$ e, portanto,

$$C_j = e^{-\theta} \frac{\theta^j}{j!}. \quad (4.22)$$

Com este resultado podemos encontrar a solução estacionária para o problema de dois sítios, quando a variável importante passa a ser a concentração do número total de mutações ($i = j + k$) que um indivíduo carrega, S_i . Deste modo,

$$\begin{aligned} S_i &= \sum_{jk} C_{jk} \delta_{j+k,i} \\ S_i &= \sum_{j \leq i} C_{ji-j} \\ S_i &= \sum_{j \leq i} C_j C_{i-j}. \end{aligned} \quad (4.23)$$

4 O limite determinístico: população infinita

Usando (4.22)

$$\begin{aligned}
 S_i &= \sum_{j \leq i} \frac{e^{-\theta} \theta^j}{j!} \frac{e^{-\theta} \theta^{i-j}}{(i-j)!} \\
 S_i &= \sum_{j \leq i} \frac{e^{-2\theta} \theta^i}{j!(i-j)!}.
 \end{aligned} \tag{4.24}$$

Multiplicando e dividindo (4.24) por $i!$,

$$S_i = \frac{e^{-2\theta} \theta^i}{i!} \sum_{m \leq i} \frac{i!}{m!(i-m)!} \tag{4.25}$$

e identificando a soma como 2^i obtemos finalmente

$$S_i = \frac{e^{-2\theta} (2\theta)^i}{i!}. \tag{4.26}$$

Esta é a probabilidade de se retirar um indivíduo com $i = j + k$ mutações da população no estado estacionário. Ela é dada por uma Poisson de média (e variância) 2θ . Note que $2\theta = \frac{2u}{s} = \frac{U}{s}$, onde U é a taxa de mutação por indivíduo.

Foi desenvolvido um programa de computador que simula a evolução temporal de C_{jk} para qualquer concentração inicial ($t = 0$) e parâmetros r, s e u . Comparando termo a termo os valores de C_{jk} entre dois tempos consecutivos, chega-se no equilíbrio para uma dada precisão previamente escolhida. Os resultados numéricos assim obtidos concordam perfeitamente com os deduzidos nesta seção. Naturalmente, a generalização para o caso de L genes é imediata, bastando substituir o fator 2 por L na fórmula (4.26).

Concluindo esse capítulo, devemos mencionar que os resultados apresentados aqui são originais e uma extensão interessante seria considerarmos relevos de replicação não fatoráveis, por exemplo, $w_k = (1 - s)^{k^\alpha}$ com $\alpha \neq 1$, para os quais esperamos que a fatoração $C_{jk} = C_j C_k$ não seja válida. No capítulo seguinte, vamos considerar o problema de população finita onde não é possível uma abordagem analítica. De fato, em tal regime não existe uma solução de equilíbrio em que as concentrações das classes (S_i) tornam-se independentes do tempo. Como veremos, para i fixo qualquer temos $S_i \rightarrow 0$ para tempos suficientemente grandes: essa é a essência da catraca de Muller.

Capítulo 5

Simulações

No princípio ($t = 0$), a população consiste de N indivíduos de L genes sem mutação alguma. À medida que o tempo evolui, podem acontecer as mutações (designadas por k), que são decorrentes da aplicação de uma Poisson para cada gene dos indivíduos. Como não existem mutações reversas, k só pode aumentar. Esse procedimento é conhecido como modelo dos infinitos alelos, pois cada gene pode receber ilimitados ks . Nele, todas as vezes que uma nova mutação é adicionada ao genótipo de um indivíduo ela nunca acontece no mesmo sítio, de modo que dois indivíduos nunca são idênticos, mesmo que possuam os mesmos ks em genes correspondentes.

Desejamos recolher dados das populações quando elas já estão em equilíbrio¹, o que nos leva a deixar correr $t = 4000$ gerações para que a primeira amostragem - de $t = 4000$ a $t = 5000$ - seja feita. A segunda e subseqüentes amostragens, por sua vez, são tomadas em outras populações que também evoluem a partir de condições iniciais idênticas até $t = 5000$ gerações. É interessante ressaltar que a população entra em equilíbrio muito rapidamente, em menos do que 1000 gerações. Dessa forma, em $t = 4000$ ela já está no equilíbrio há, pelo menos, 3000 gerações, tempo considerável para que seja desprezado qualquer vestígio de comportamento transiente.

¹Aqui, entende-se por equilíbrio o regime estacionário em que as classes de indivíduos são perdidas a uma taxa constante, isto é, os gráficos de grandezas associadas ao número de mutações na população contra o tempo são retas.

5 Simulações

Para cada indivíduo das populações é calculado o número de mutações totais que ele carrega por geração², que já designamos k . São gravados, por geração, o menor e maior valores de k , k_{min} e k_{max} , respectivamente, bem como a média de k , k_{medio} . São tiradas médias desses valores em diversos tamanhos de populações (designados $NPOP$): $NPOP = 50, 100$ ou 500 e essas novas variáveis são designadas por $\langle k_{min} \rangle$, $\langle k_{max} \rangle$ e $\langle k_{medio} \rangle$. A velocidade da catraca de Muller é obtida através da inclinação do ajuste das curvas de $\langle k_{min} \rangle \times t$ no equilíbrio.

Para tamanho de população $N = 100$, probabilidade máxima de recombinação, $r = 1$, e $NPOP = 100$, variamos o tamanho L dos indivíduos a fim de verificarmos a dependência de $\langle k_{min} \rangle$ em função das gerações. Consideramos as duas dinâmicas descritas no Capítulo 2: na Figura 5.1, a Dinâmica 1 e, na Figura 5.2, a Dinâmica 2. Em ambas, L varia de 2 a 1000 em intervalos não regulares ($L = 2, 3, 4, 5, 10, 20, 50, 100, 500$ e 1000), sendo a taxa de mutação por indivíduo $U = \frac{0.5}{L}$ à esquerda e $U = \frac{1.0}{L}$ à direita³, com $s = 0.1$ na parte superior e $s = 0.01$ na parte inferior. Para as duas dinâmicas, dado um t fixo, observamos o decréscimo de $\langle k_{min} \rangle$ com o crescimento de L , o que provém do fato de que apesar da recombinação formar indivíduos com ks grandes (através da união de gametas com muitas mutações) ela também forma indivíduos com poucas mutações. Numa posterior ação da seleção são eliminados aqueles indivíduos que carregam muitas mutações, sendo mantidos aqueles com $\langle k_{min} \rangle$ pequeno, isto é, os mais adaptados. Ainda nas Figuras 5.1 e 5.2, quanto maior a taxa de mutação U , maior é o valor de $\langle k_{min} \rangle$ alcançado na última geração, $\langle k_{min}(t = 5000) \rangle$. O mesmo se verifica se s é pequeno: quanto menor o seu valor, menores são as taxas de rejeição da população às mutações, elevando $\langle k_{min} \rangle$ [22].

A velocidade R da catraca relativa às Figuras 5.1 e 5.2 pode ser vista na Tabela 5.3, com $N = 100$. Outros valores de N , como 10, 50 e 200 são encontrados nas Tabelas 5.1, 5.2 e 5.4, respectivamente. Em todas elas foram acrescentadas duas colunas com $U = \frac{5.0}{L}$, uma com $s = 0.01$

²Lembrar que esse número é resultado da soma das mutações em todos os genes do indivíduo.

³Note que a taxa de mutação por gene é, então, constante, independente de L .

5 Simulações

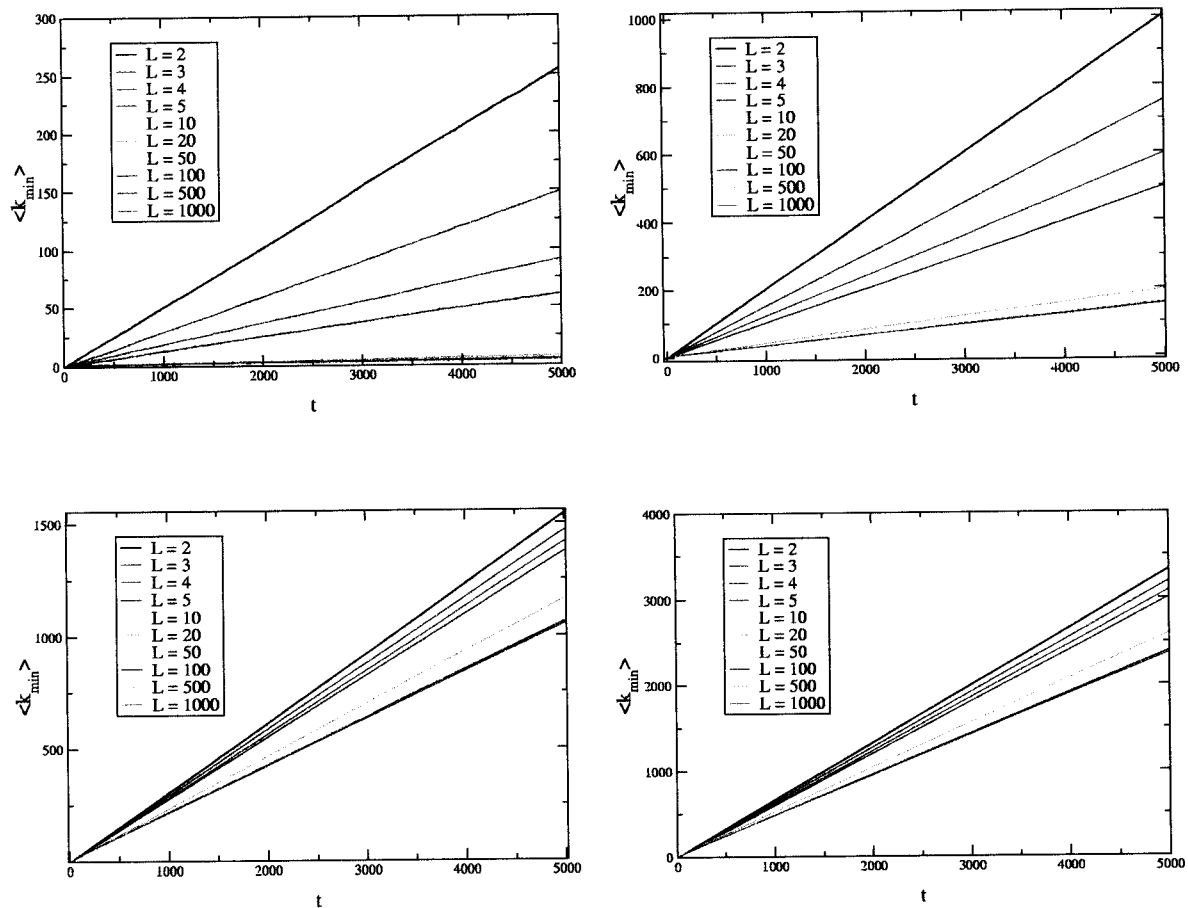


Figura 5.1: Dinâmica 1: número mínimo de mutações por indivíduo por geração mediados em 100 populações de $N = 100$ indivíduos para $s = 0.1$ (superiores) e $s = 0.01$ (inferiores), com $r = 1.0$. À esquerda, $U = \frac{0.5}{L}$, e à direita, $U = \frac{1.0}{L}$.

5 Simulações

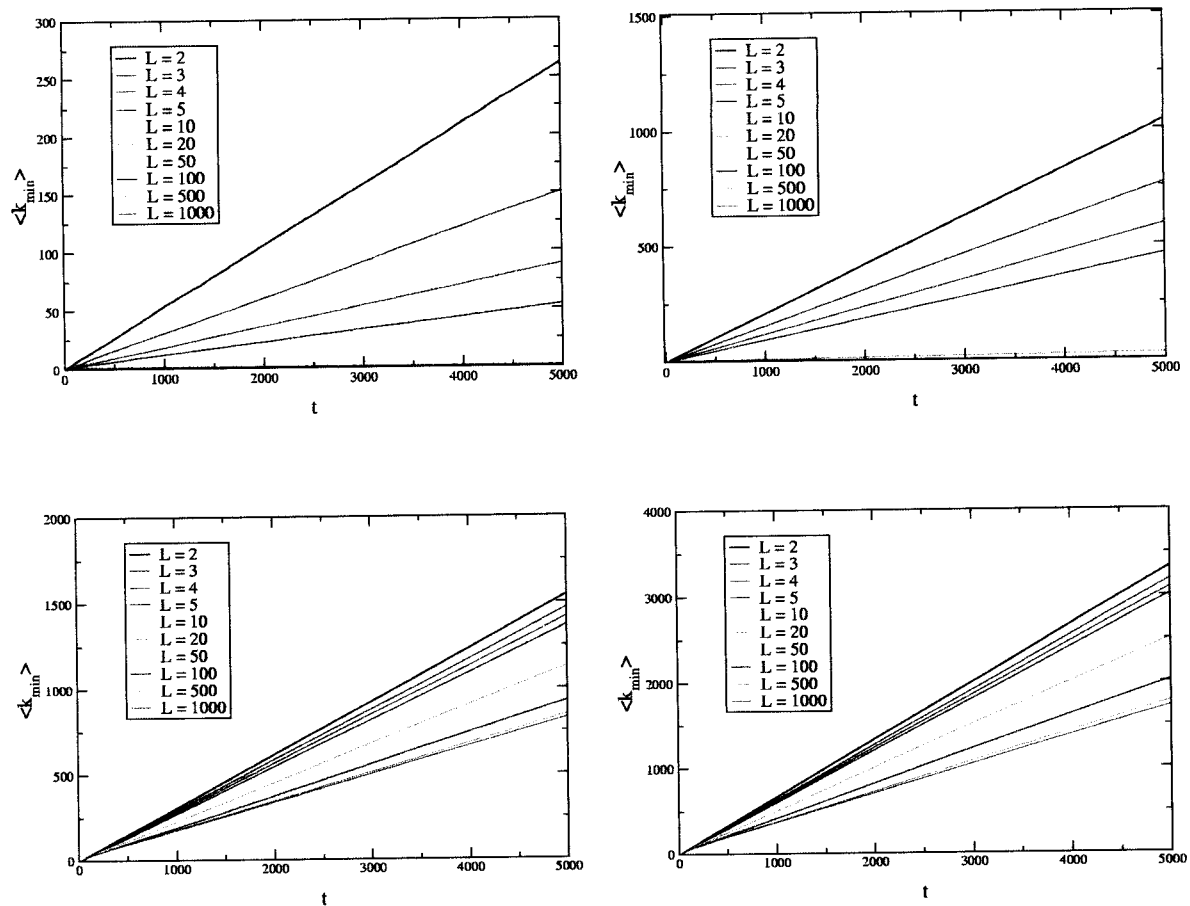


Figura 5.2: Dinâmica 2: número mínimo de mutações por indivíduo por geração mediados em 100 populações de $N = 100$ indivíduos para $s = 0.1$ (superiores) e $s = 0.01$ (inferiores), com $r = 1.0$. À esquerda, $U = \frac{0.5}{L}$, e à direita, $U = \frac{1.0}{L}$.

5 Simulações

e outra com $s = 0.1$. Como era de se esperar, R (para um dado valor de mutação) é maior se o parâmetro seletivo é pequeno, decrescendo com o acréscimo de L , mantendo-se s fixo. Ainda com s fixo, a velocidade da catraca cresce bastante com U . Nos casos simulados, esse crescimento de R é linear quando $N = 10$ e $s = 0.01$, mas não acontece para $N = 50, 100$ e 200 , quaisquer sejam os valores de s . Em todas as tabelas, o maior valor da velocidade da catraca é aquele para o qual $U = \frac{5.0}{L}$ e $s = 0.01$, isto é, taxa de mutação máxima e dano mínimo.

No geral, R não varia muito dentro de um mesmo s para populações pequenas. Isto não é verdade se N cresce, pois quanto maior N , maior a variação em R . Se a taxa de mutação é pequena, $U = \frac{0.5}{L}$, mas cada mutação está associada a uma grande perda, a velocidade da catraca decai rapidamente, o mesmo sendo observado se $U = \frac{1.0}{L}$. A Dinâmica 2 sempre proporciona as maiores quedas nos valores de R , pois se existe maior variedade de filhos, é natural que a seleção atue em prol dos mais aptos. Quanto maior L , a variedade de filhos que podem ser gerados é ainda maior, o que leva a uma queda considerável nos valores de $\langle k_{min} \rangle$ e, conseqüentemente, de R .

Para melhor visualização, todas as colunas das tabelas estão expostas em gráficos e podem ser encontradas nas Figuras 5.3 (Dinâmica 1) e 5.4 (Dinâmica 2). Num mesmo gráfico estão os valores de R em função de L para $N = 10, 50, 100$ e 200 . Se lidas da esquerda para a direita e de cima para baixo, na primeira "linha" das Figuras está $U = \frac{0.5}{L}$, com $s = 0.01$ na primeira coluna e $s = 0.1$ na segunda coluna. Na segunda linha está $U = \frac{1.0}{L}$ e na terceira, $U = \frac{5.0}{L}$. Em ambas as figuras, notamos que somente para $N = 10$ e $s = 0.01$ R se mantém constante, enquanto que para todas as outras combinações de valores R sempre decresce, e tanto mais quanto maior for o valor de N , tendendo a zero para praticamente todos os casos em que $s = 0.1$, menos para $U = \frac{5.0}{L}$. É visível, assim, que a catraca pode ser parada, desde que haja uma combinação correta de parâmetros para as dinâmicas estudadas.

Se aumentamos o tamanho da população e mantemos todos os outros parâmetros fixos, notamos que a velocidade da catraca diminui com o acréscimo de N , como pode ser visto na Tabela

5 Simulações

Velocidade da Catraca para $N=10$						
Dinâmica 1						
L	U=0.5/L		U=1.0/L		U=5.0/L	
	s=0.01	s=0.1	s=0.01	s=0.1	s=0.01	s=0.1
1	0.455993	0.264725	0.913058	0.583771	4.625329	3.553344
2	0.460129	0.237981	0.918579	0.532265	4.619474	3.328737
3	0.459063	0.228813	0.917118	0.511081	4.601253	3.243833
4	0.458471	0.223862	0.913936	0.501686	4.586714	3.204289
5	0.459746	0.224284	0.918823	0.496327	4.595142	3.200811
10	0.453027	0.213994	0.910041	0.483878	4.568173	3.149960
20	0.456352	0.214757	0.913493	0.483498	4.576359	3.147518
50	0.459828	0.214161	0.914285	0.481556	4.579908	3.156625
100	0.453827	0.213555	0.916612	0.482909	4.568928	3.168400
200	0.455326	0.210659	0.913587	0.481888	4.580159	3.159298
300	0.455680	0.215390	0.916739	0.482768	4.576204	3.159384
400	0.454968	0.216063	0.909001	0.483598	4.589855	3.169252
500	0.457386	0.215391	0.915206	0.482980	4.586695	3.162232
1000	0.456619	0.215460	0.909335	0.478779	4.581396	3.159781
Dinâmica 2						
1	0.458836	0.263177	0.917886	0.583657	4.641316	3.548462
2	0.456723	0.240311	0.917469	0.536686	4.601780	3.368311
3	0.453176	0.229495	0.912493	0.514892	4.594941	3.278104
4	0.456660	0.224067	0.916122	0.496511	4.587821	3.201846
5	0.453380	0.218700	0.909618	0.485783	4.592793	3.146954
10	0.458039	0.204842	0.916200	0.459302	4.577853	3.022684
20	0.456340	0.195741	0.916545	0.435575	4.565066	2.920629
50	0.456054	0.189842	0.916246	0.420077	4.574047	2.826624
100	0.456669	0.187678	0.909730	0.407695	4.567645	2.786711
200	0.453313	0.183148	0.911488	0.403463	4.573165	2.764581
300	0.457408	0.184454	0.904387	0.405003	4.553299	2.755451
400	0.461022	0.183514	0.907695	0.405102	4.574148	2.756645
500	0.456898	0.185741	0.908591	0.403573	4.558520	2.751546
1000	0.456059	0.186439	0.905845	0.401316	4.553717	2.745685

Tabela 5.1: Velocidade da Catraca de Muller em função de L e U para $r = 1.0$ e $N = 10$ mediados em $NPOP = 100$ populações. Acima, Dinâmica 1. Abaixo, Dinâmica 2.

5 Simulações

Velocidade da Catraca para $N=50$						
Dinâmica 1						
L	U=0.5/L		U=1.0/L		U=5.0/L	
	s=0.01	s=0.1	s=0.01	s=0.1	s=0.01	s=0.1
1	0.367735	0.134612	0.765941	0.368169	4.183833	2.874591
2	0.351311	0.089218	0.733970	0.272797	4.052202	2.494259
3	0.341832	0.066119	0.718540	0.227528	3.969585	2.313939
4	0.339649	0.052579	0.704406	0.199190	3.904006	2.205637
5	0.334864	0.045370	0.699493	0.177737	3.867096	2.134318
10	0.324804	0.027411	0.679183	0.138281	3.749642	2.001114
20	0.316557	0.021086	0.656796	0.117928	3.667000	1.940929
50	0.308445	0.018740	0.643574	0.111947	3.622005	1.931739
100	0.310395	0.018084	0.639560	0.111008	3.602895	1.934471
200	0.310411	0.018588	0.643890	0.111492	3.600264	1.930662
300	0.311773	0.018658	0.642525	0.111703	3.614486	1.932176
400	0.308105	0.018012	0.639422	0.111117	3.601381	1.939441
500	0.309075	0.018288	0.645996	0.110291	3.603534	1.936804
1000	0.308235	0.018653	0.638255	0.111483	3.607106	1.934341
Dinâmica 2						
1	0.364998	0.136088	0.769997	0.369777	4.181872	2.878669
2	0.354539	0.091815	0.742523	0.276434	4.051401	2.540934
3	0.343779	0.067123	0.717729	0.230049	3.963961	2.339240
4	0.337540	0.051558	0.705985	0.195442	3.912072	2.198176
5	0.332592	0.042246	0.694241	0.171204	3.847935	2.083234
10	0.322571	0.017227	0.668639	0.100492	3.713306	1.748622
20	0.311634	0.006806	0.642457	0.050796	3.558766	1.446778
50	0.305810	0.002238	0.619951	0.019099	3.391711	1.112290
100	0.299872	0.001133	0.604929	0.010863	3.265529	0.933651
200	0.298437	0.000925	0.603385	0.006434	3.187877	0.806580
300	0.294638	0.000826	0.598470	0.005561	3.154181	0.756012
400	0.292765	0.000808	0.595658	0.005601	3.135284	0.725174
500	0.296544	0.000831	0.594836	0.005290	3.118142	0.711624
1000	0.295706	0.000788	0.592829	0.004783	3.081438	0.672861

Tabela 5.2: Velocidade da Catraca de Muller em função de L e U para $r = 1.0$ e $N = 50$ mediados em $NPOP = 100$ populações. Acima, Dinâmica 1. Abaixo, Dinâmica 2.

5 Simulações

Velocidade da Catraca para $N=100$						
Dinâmica 1						
L	U=0.5/L		U=1.0/L		U=5.0/L	
	s=0.01	s=0.1	s=0.01	s=0.1	s=0.01	s=0.1
1	0.333524	0.099710	0.716168	0.303968	4.033591	2.657314
2	0.310797	0.051356	0.666120	0.201570	3.840457	2.233058
3	0.294174	0.028924	0.638870	0.150190	3.734713	2.013441
4	0.283184	0.017694	0.618160	0.119750	3.653002	1.886235
5	0.274067	0.012184	0.601810	0.100290	3.591970	1.812018
10	0.252200	0.003149	0.557400	0.056325	3.409966	1.626398
20	0.229274	0.001247	0.517240	0.037806	3.261124	1.539587
50	0.217687	0.000885	0.487040	0.031147	3.146429	1.512875
100	0.210072	0.000740*	0.475740	0.030392*	3.105372	1.510640*
200	0.210101	0.000740*	0.472168	0.030544*	3.101774	1.512914*
300	0.208776	0.000794*	0.472440	0.030741*	3.098904	1.509487*
400	0.209188	0.000721*	0.468162	0.030197*	3.098049	1.511466*
500	0.207568	0.000788*	0.470660	0.030180*	3.102446	1.511955*
1000	0.207357	0.000731	0.472276	0.030266	3.095049	1.508589*
Dinâmica 2						
1	0.334725	0.097386	0.712592	0.302001	4.026806	2.659908
2	0.310210	0.052141	0.666264	0.206818	3.855887	2.276032
3	0.296359	0.029338	0.639638	0.152490	3.730304	2.040018
4	0.283638	0.017518	0.617925	0.116894	3.655354	1.880020
5	0.274057	0.010520	0.603374	0.090204	3.590500	1.750315
10	0.249716	0.001115	0.552468	0.029910	3.355398	1.362251
20	0.222997	0.000040	0.493637	0.004888	3.118210	1.004228
50	0.198767	0.000000	0.439812	0.000363	2.799986	0.613114
100	0.183603	0.000000	0.398994	0.000137	2.564842	0.406060
200	0.176420	0.000000	0.369860	0.000016	2.357798	0.268070
300	0.170653	0.000025	0.356779	0.000122	2.238693	0.221223
400	0.164752	0.000000	0.355241	0.000000	2.184333	0.195210
500	0.166644	0.000008	0.347304	0.000014	2.126792	0.180198
1000	0.163126	0.000000	0.336423	0.000064	2.009410	0.153950

Tabela 5.3: Velocidade da Catraca de Muller em função de L e U para $r = 1.0$ e $N = 100$ mediados em $NPOP = 100$ populações. Para os termos destacados com (*), $NPOP = 500$.

5 Simulações

Velocidade da Catraca para $N=200$						
Dinâmica 1						
L	U=0.5/L		U=1.0/L		U=5.0/L	
	s=0.01	s=0.1	s=0.01	s=0.1	s=0.01	s=0.1
1	0.303049	0.071100	0.670225	0.249918	3.899274	2.467395
2	0.271495	0.025653	0.608724	0.145135	3.688424	2.002227
3	0.248934	0.008396	0.570481	0.092755	3.542164	1.761101
4	0.237825	0.002999	0.546505	0.063316	3.449162	1.626060
5	0.225803	0.000771	0.525747	0.044389	3.358680	1.534774
10	0.191963	0.000003	0.457009	0.013480	3.119858	1.318315
20	0.161541	0.000022	0.397574	0.004312	2.909356	1.209803
50	0.130408	0.000000	0.345225	0.002235	2.720908	1.151731
100	0.119462	0.000000	0.321329	0.002289	2.648667	1.148039
200	0.114838	0.000021	0.309540	0.002406	2.619652	1.145614
300	0.112511	0.000000	0.305331	0.001937	2.618037	1.142992
400	0.110305	0.000000	0.307317	0.002037	2.621845	1.146937
500	0.110628	0.000000	0.303636	0.002274	2.616581	1.149592
1000	0.111697	0.000000	0.305717	0.002453	2.609826	1.143980
Dinâmica 2						
1	0.302535	0.071020	0.666863	0.248818	3.891566	2.464026
2	0.271507	0.025783	0.604318	0.149572	3.685439	2.046273
3	0.254563	0.009271	0.573462	0.094352	3.547753	1.798389
4	0.236920	0.002444	0.542996	0.061799	3.441955	1.611059
5	0.224516	0.000562	0.524638	0.038499	3.357433	1.470135
10	0.190271	0.000000	0.453752	0.002988	3.069746	1.042497
20	0.154015	0.000000	0.380032	0.000000	2.749223	0.666369
50	0.109345	0.000010	0.286138	0.000006	2.304408	0.265873
100	0.083106	0.000000	0.224818	0.000004	1.966120	0.098971
200	0.067018	0.000001	0.174347	0.000000	1.637339	0.032716
300	0.057938	0.000000	0.149324	0.000000	1.456259	0.018435
400	0.054774	0.000000	0.135912	0.000000	1.333822	0.012936
500	0.051533	0.000000	0.126082	0.000000	1.243604	0.010848
1000	0.046662	0.000000	0.105634	0.000085	1.013012	0.005345

Tabela 5.4: Velocidade da Catraca de Muller em função de L e U para $r = 1.0$ e $N = 200$ mediados em $NPOP = 100$ populações. Acima, Dinâmica 1. Abaixo, Dinâmica 2.

5 Simulações

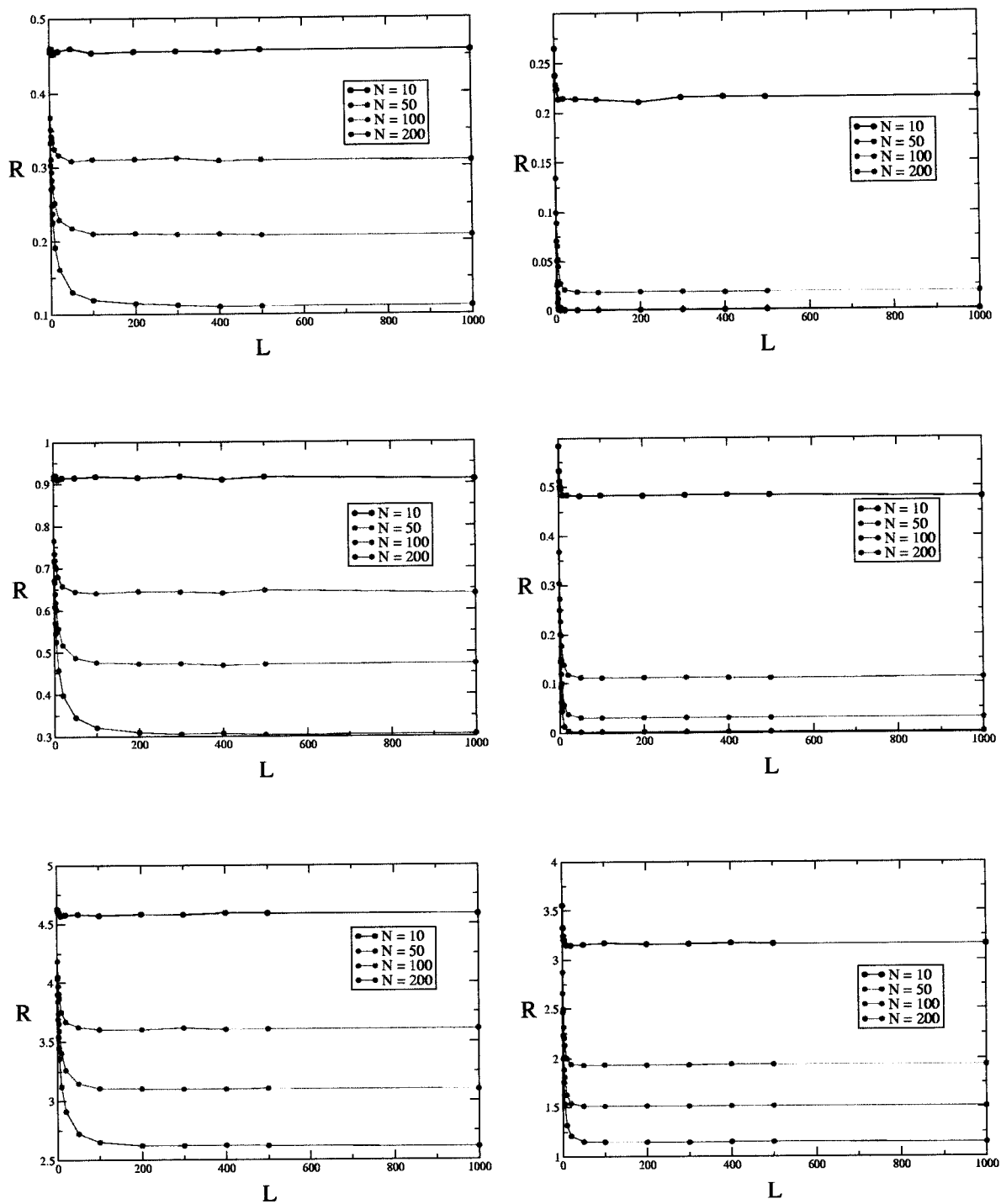


Figura 5.3: Dinâmica 1: R em função de L para $s = 0.01$ (à esquerda) e $s = 0.1$ (à direita). Na parte superior, $U = \frac{0.5}{L}$; na intermediária, $U = \frac{1.0}{L}$; e na inferior, $U = \frac{5.0}{L}$.

5 Simulações

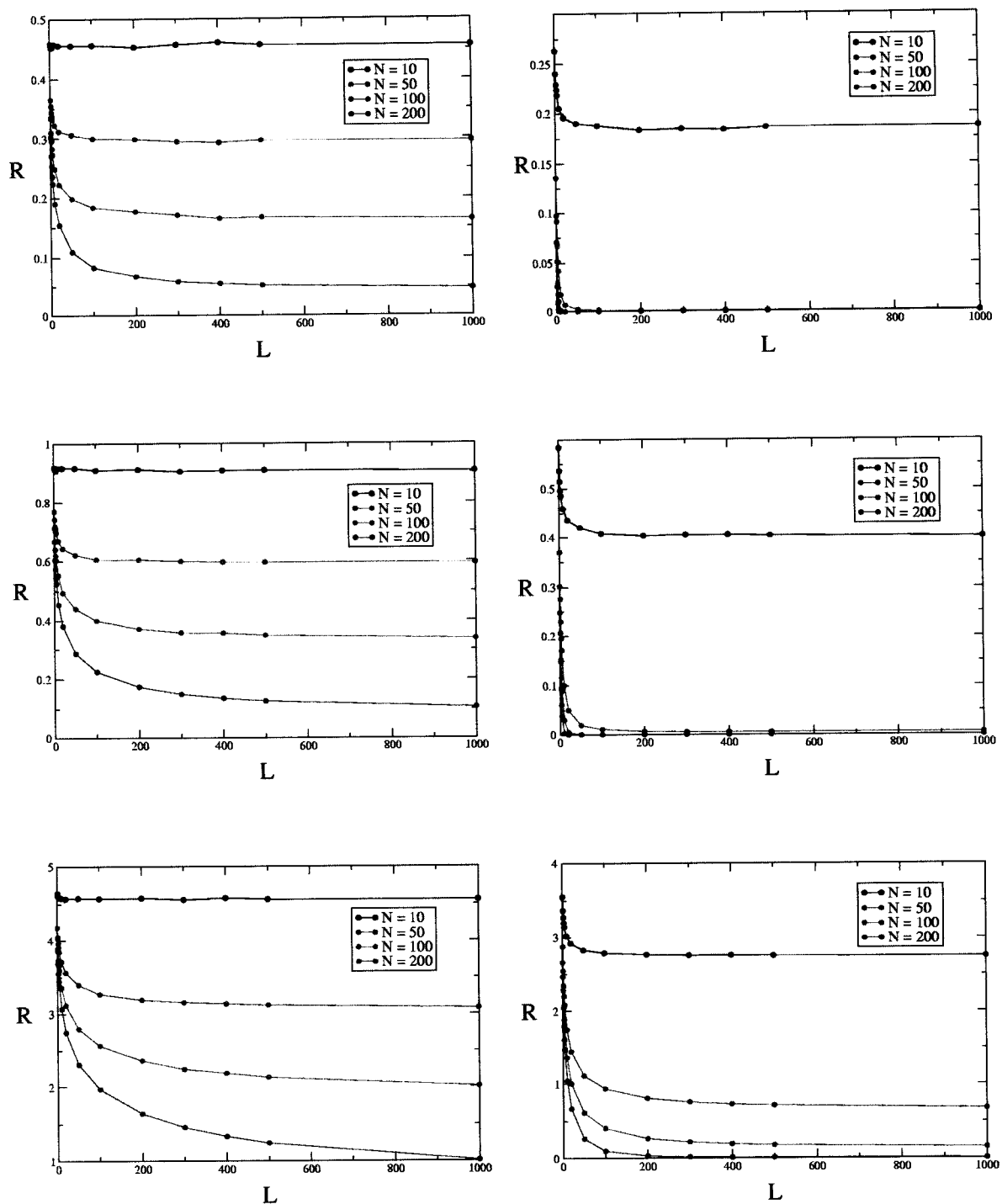


Figura 5.4: Dinâmica 2: R em função de L para $s = 0.01$ (à esquerda) e $s = 0.1$ (à direita). Na parte superior, $U = \frac{0.5}{L}$; na intermediária, $U = \frac{1.0}{L}$; e na inferior, $U = \frac{5.0}{L}$.

5 Simulações

5.5. Novamente, R atinge o mínimo quando $U = \frac{0.5}{L}$ e $s = 0.1$.

Velocidade da Catraca para N variando e $L=2$

N	Dinâmica 1					
	U=0.5/L		U=1.0/L		U=5.0/L	
	s=0.01	s=0.1	s=0.01	s=0.1	s=0.01	s=0.1
10	0.460129	0.237981	0.918579	0.532265	4.619474	3.328737
50	0.351311	0.089218	0.733970	0.272797	4.052202	2.494259
100	0.306409	0.051241	0.656704	0.199660	3.840457	2.233058
200	0.271495	0.025653	0.608724	0.145135	3.688424	2.002227
300	0.253102	0.014719	0.582829	0.117490	3.611489	1.885201
400	0.242164	0.008807	0.566106	0.105359	3.524296	1.811381
500	0.225887	0.006368	0.543575	0.087050	3.492640	1.744287
600	0.226662	0.003944	0.531808	0.075936	3.470334	1.702585
700	0.218785	0.002033	0.525254	0.072696	3.416822	1.673785
800	0.210263	0.002224	0.520840	0.066908	3.426630	1.631735
900	0.213129	0.001019	0.511535	0.062047	3.374732	1.617787
1000	0.207771	0.000708	0.502927	0.058523	3.372990	1.598481
2000	0.184804	0.000000	0.463142	0.035192	3.243652	1.443502
3000	0.178796	0.000000	0.449256	0.021957	3.213292	1.369243
4000	0.171499	0.000000	0.433494	0.016703	3.153955	1.317790
5000	0.162309	0.000000	0.435100	0.013482	3.122468	1.274190

Tabela 5.5: Velocidade da Catraca de Muller em função de N , U e s para $r = 1.0$ e $L = 2$ mediados em $NPOP = 10$ populações. Dados apenas da Dinâmica 1.

Também plotamos como a velocidade da catraca depende do tamanho da população quando os indivíduos possuem somente um gene, ou seja, $L = 1$, e o parâmetro seletivo vale $s = 0.1$, como pode ser visto na Figura 5.5.

Notamos que quanto menores as taxas de mutação, mais rapidamente a catraca tende a zero, não sendo preciso avançar para grandes tamanhos de população. Para os maiores valores de U , entretanto, a tendência à retenção da catraca parece ser lenta.

Para $N = 100$, parâmetro seletivo $s = 0.1$, taxa de mutação por gene $U = \frac{0.5}{L}$ e $NPOP = 100$,

5 Simulações

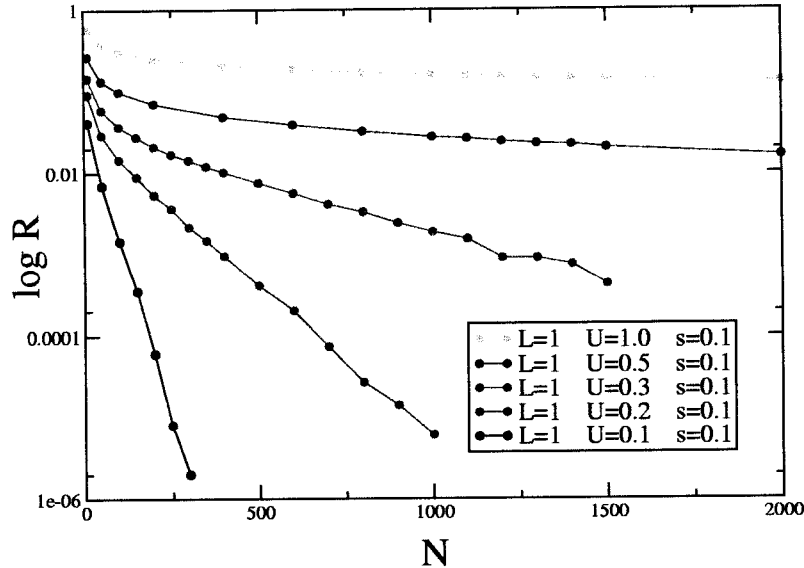


Figura 5.5: Logaritmo da velocidade da catraca em função do tamanho N da população para $L = 1$ e $r = 1.0$. São variados os valores da taxa de mutação, U , e s é mantido fixo em 0.1.

com $L = 2$ e 1000 variamos a taxa de recombinação r e plotamos a dependência de $\langle k_{min} \rangle$ com t , como mostram as Figuras 5.6 e 5.7. Notamos que, para uma dada geração t , $\langle k_{min} \rangle$ decresce com r crescente, ou seja, o mecanismo de manter L fixo e variar r produz o mesmo efeito do mecanismo de se fixar r e variar L , mas, neste caso, porque a recombinação está ajudando a criar na população indivíduos com poucas mutações. É natural que quanto maior a probabilidade de haver recombinação, menor seja o número de indivíduos com muitas mutações mantidos de uma geração para outra. As Figuras 5.8 e 5.9, relativas à Dinâmica 2, mostram resultados semelhantes.

Como podemos perceber, quando passamos de parâmetro seletivo $s = 0.1$ para $s = 0.01$ (mantendo U fixo), em ambas as Dinâmicas $\langle k_{min} \rangle$ aumenta consideravelmente e não faz muita diferença, para $L = 2$, haver recombinação ou não, pois em geral, todos os indivíduos - sejam eles recombinantes ou não - possuem um número bem próximo de mutações nos seus dois genes. Quando $L = 1000$, todavia, as mutações podem estar mais distribuídas, levando a uma

5 Simulações

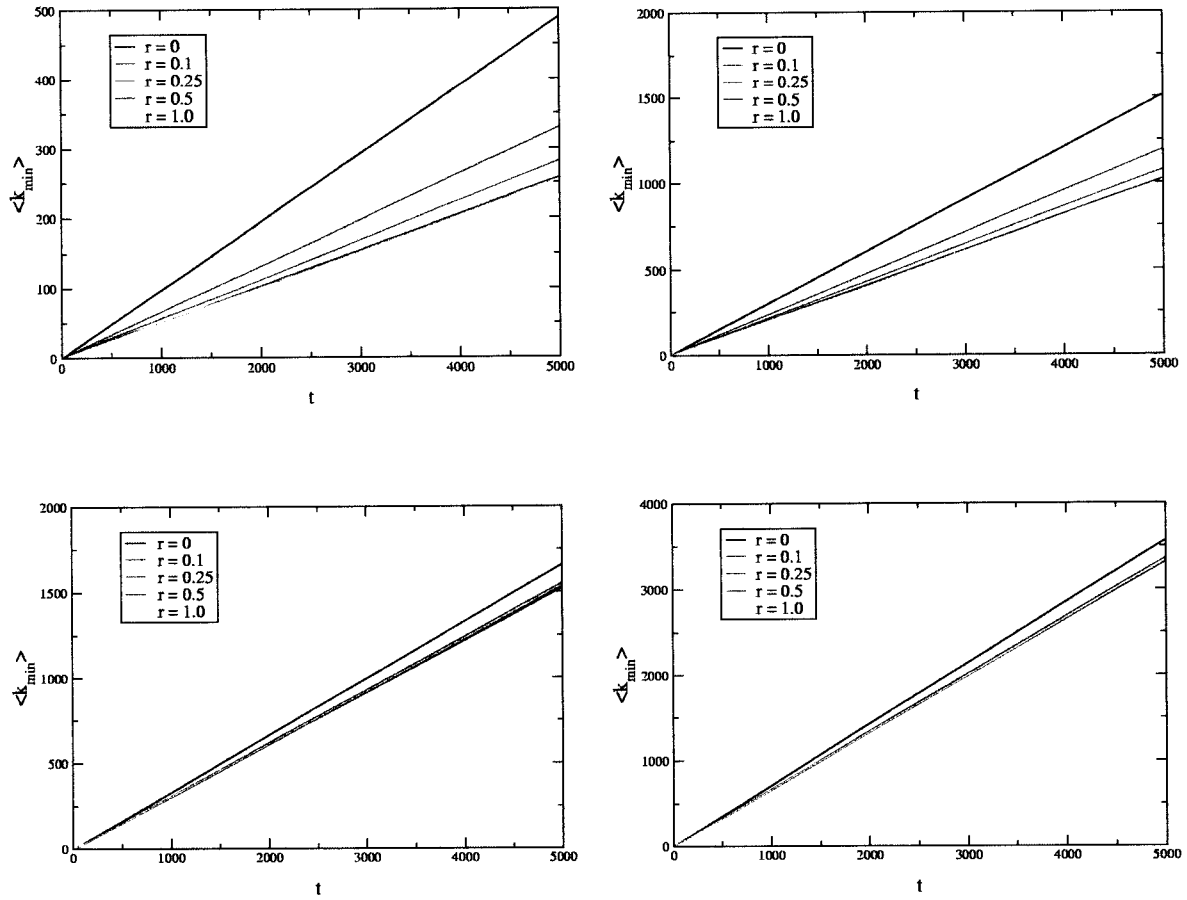


Figura 5.6: Dinâmica 1: número mínimo de mutações por indivíduo em função das gerações mediados em 100 populações de $N = 100$ indivíduos, $L = 2$ e r variando. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$. À esquerda, $U = \frac{0.5}{L}$, e à direita, $U = \frac{1.0}{L}$.

5 Simulações

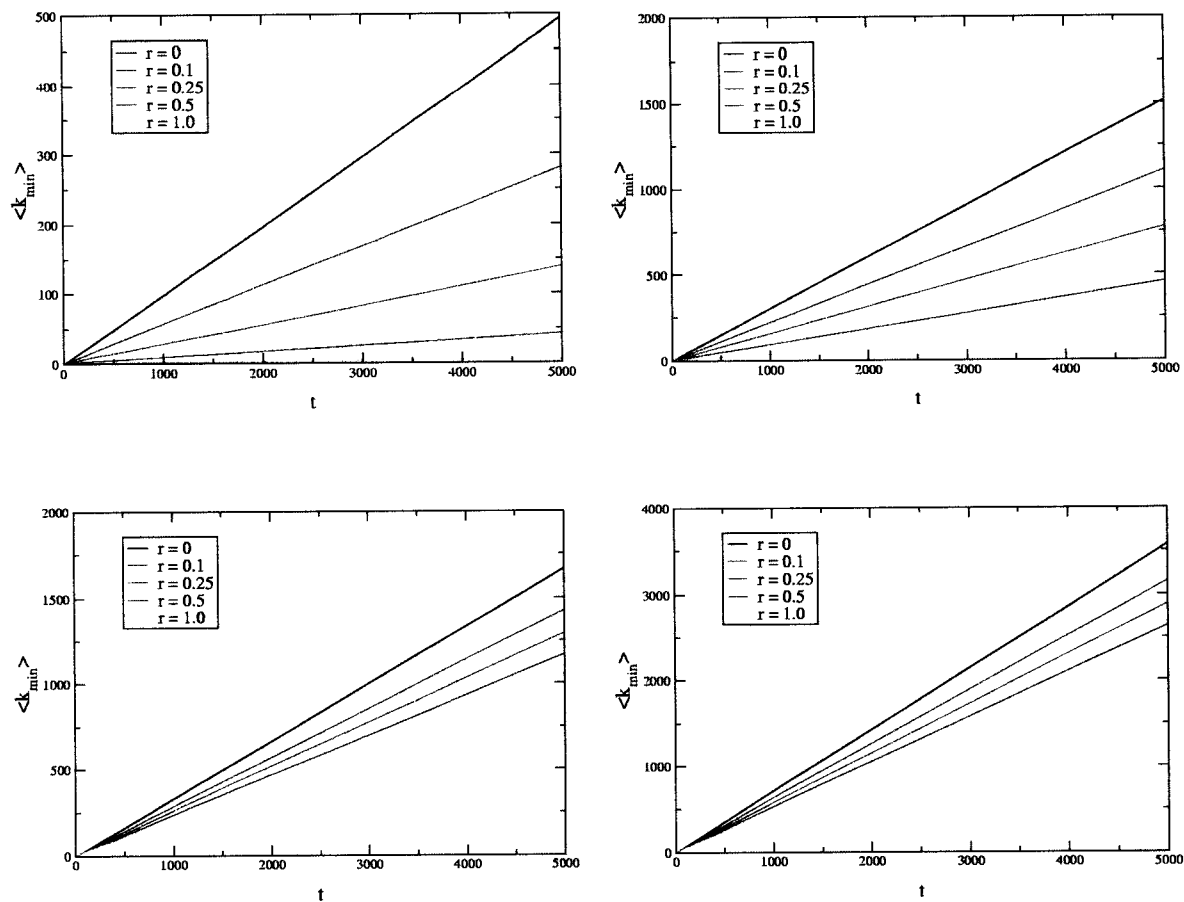


Figura 5.7: Dinâmica 1: número mínimo de mutações por indivíduo em função das gerações mediados em 100 populações de $N = 100$ indivíduos, $L = 1000$ e r variando. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$. À esquerda, $U = \frac{0.5}{L}$, e à direita, $U = \frac{1.0}{L}$.

5 Simulações

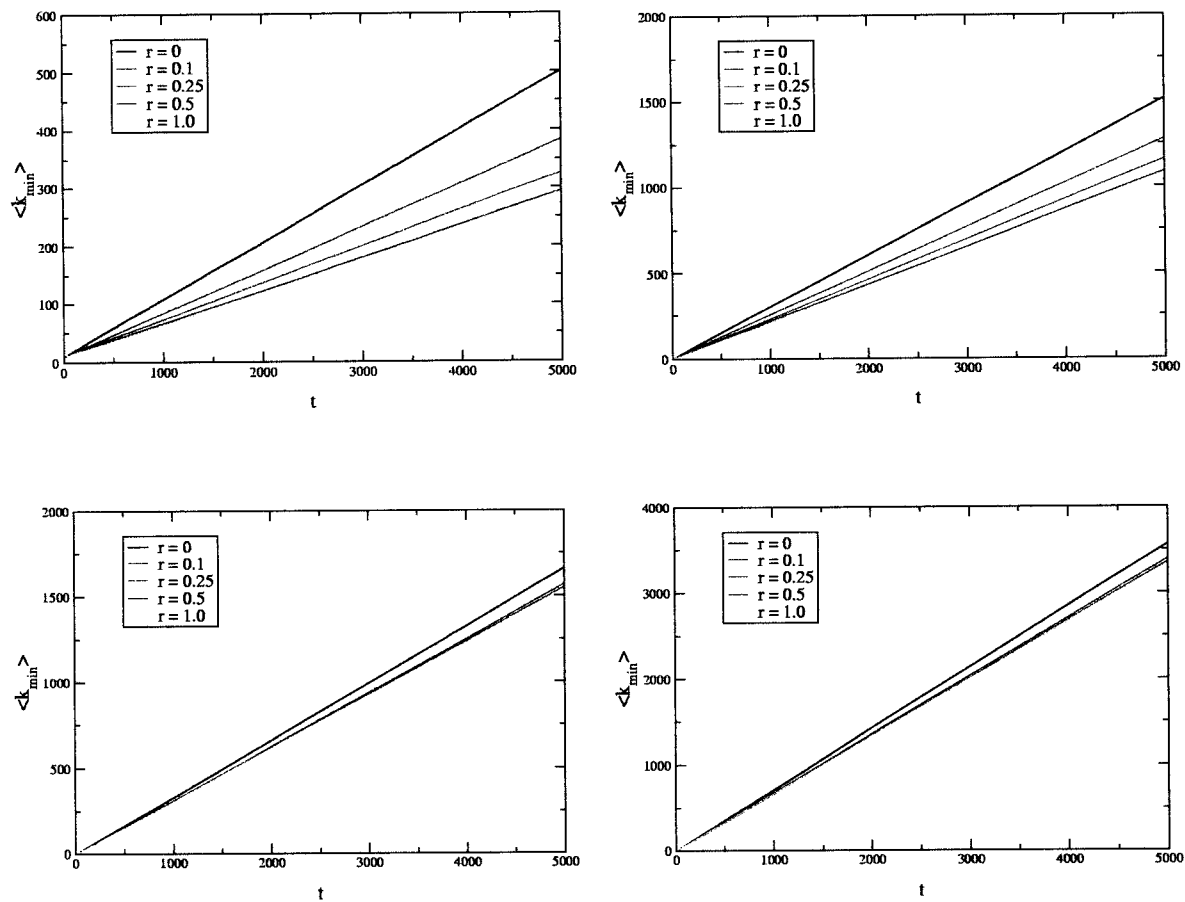


Figura 5.8: Dinâmica 2: número mínimo de mutações por indivíduo em função das gerações mediados em 100 populações de $N = 100$ indivíduos, $L = 2$ e r variando. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$. À esquerda, $U = \frac{0.5}{L}$, e à direita, $U = \frac{1.0}{L}$.

5 Simulações

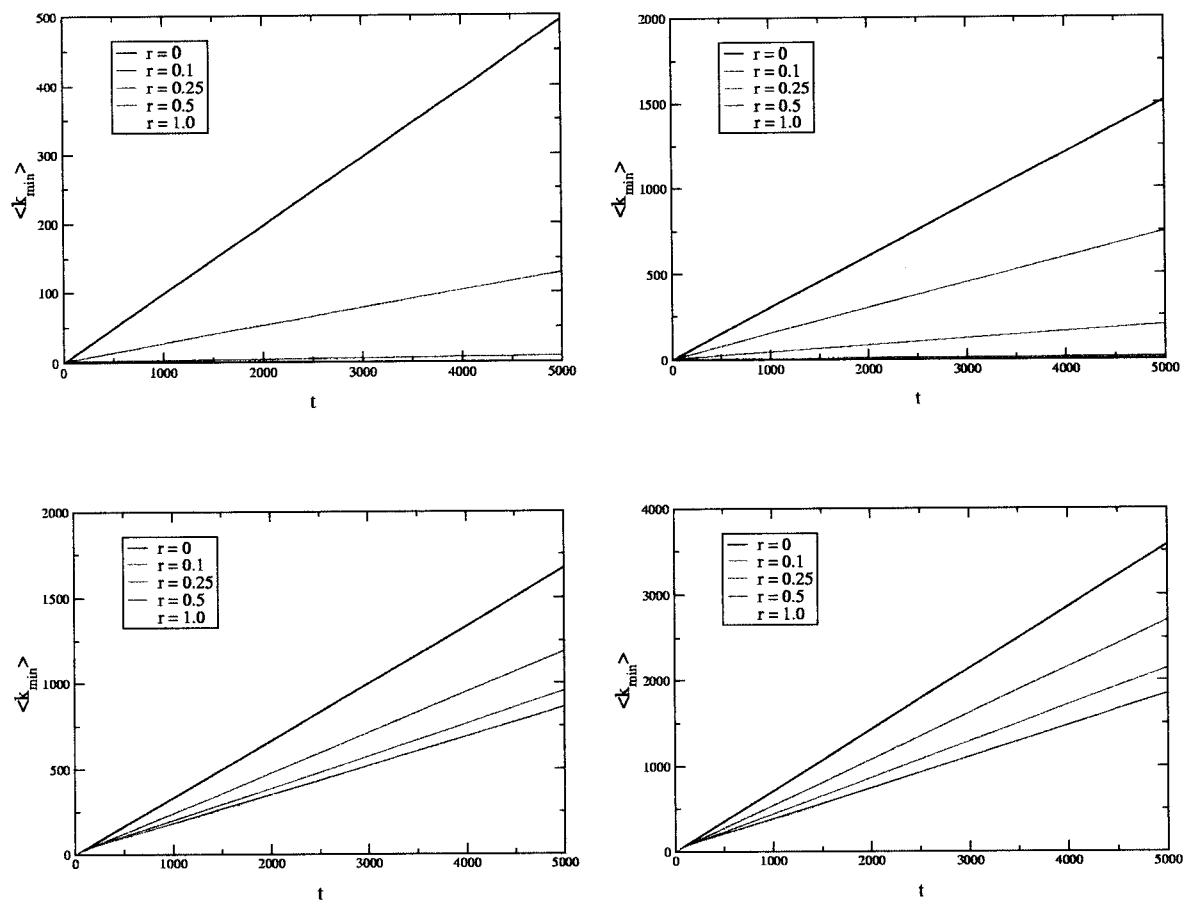


Figura 5.9: Dinâmica 2: número mínimo de mutações por indivíduo em função das gerações mediados em 100 populações de $N = 100$ indivíduos, $L = 1000$ e r variando. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$. À esquerda, $U = \frac{0.5}{L}$, e à direita, $U = \frac{1.0}{L}$.

5 Simulações

maior probabilidade de agrupamentos de genes com poucas mutações se r for cada vez maior. Da mesma forma que para $L = 2$, se o grau deletério das mutações for menor, os indivíduos as carregarão em maior quantidade. Resultado semelhante é encontrado se s for mantido fixo e U variado. À medida que a taxa de mutação aumenta, $\langle k_{min} \rangle$ a acompanha. Também é interessante notar que como a Dinâmica 2 implica numa maior variedade filial, o valor de $\langle k_{min} \rangle$ atingido na última geração é bem menor do que aquele para a Dinâmica 1, quando há recombinação. Na sua ausência, entretanto, torna-se independente da dinâmica estudada.

Plotando $R \times r$ para diversos tamanhos de L e s fixo em 0.1 e 0.01, com $U = \frac{0.5}{L}$ e $U = \frac{1.0}{L}$, obtemos as Figuras 5.10 (Dinâmica 1) e 5.11 (Dinâmica 2). Verificamos que a velocidade da catraca é sensível ao tamanho do indivíduo e que ela é função decrescente de r : quanto maiores r e L , menor a sua velocidade, tendendo a zero em alguns casos, em especial aqueles para os quais o parâmetro seletivo é $s = 0.1$.

Para $r = 0.1, 0.25, 0.5$ e 1.0 plotamos R em função de $\frac{1}{L}$, com $s = 0.1$ e 0.01 , e $U = \frac{0.5}{L}$ e $\frac{1.0}{L}$, como mostram as Figuras 5.12 (Dinâmica 1) e 5.13 (Dinâmica 2). Os valores de R para $r = 0$ não aparecem no gráfico pois todas as curvas foram normalizadas tomando-os por base. Notamos que aumentando L a curva de R tende a um valor de equilíbrio (se analisada da direita para a esquerda) quando trabalhamos com a primeira dinâmica e $s = 0.1$. Em particular, quando $U = \frac{0.5}{L}$ e $r = 1.0$ este equilíbrio pára a movimentação da catraca ($R \rightarrow 0$). para $U = \frac{1.0}{L}$ para que pudesse ser feita uma extrapolação no limite de L grande ($L \rightarrow \infty$). Com os resultados dos ajustes para este caso, notamos que a velocidade da catraca realmente vai a um valor de equilíbrio, mas ele é não nulo para $U = \frac{0.5}{L}$ e $r = 1.0$, apesar de ser bastante pequeno. Assim, a catraca não mais altera seu movimento quando atinge esse valor, mas também não deixa de atuar, existindo em pequena escala. O mesmo não podemos dizer quando plotamos $R \times r$ para $s = 0.01$ e $U = \frac{0.5}{L}$ e $\frac{1.0}{L}$, gráficos inferiores da Figura 5.12. Existe uma grande flutuação nos resultados e até mesmo as concavidades das curvas são diferentes daquelas obtidas quando $s = 0.1$. Ainda conseguimos perceber que os menores valores de R ocorrem para os maiores valores de L , embora não mais consigamos observar uma

5 Simulações

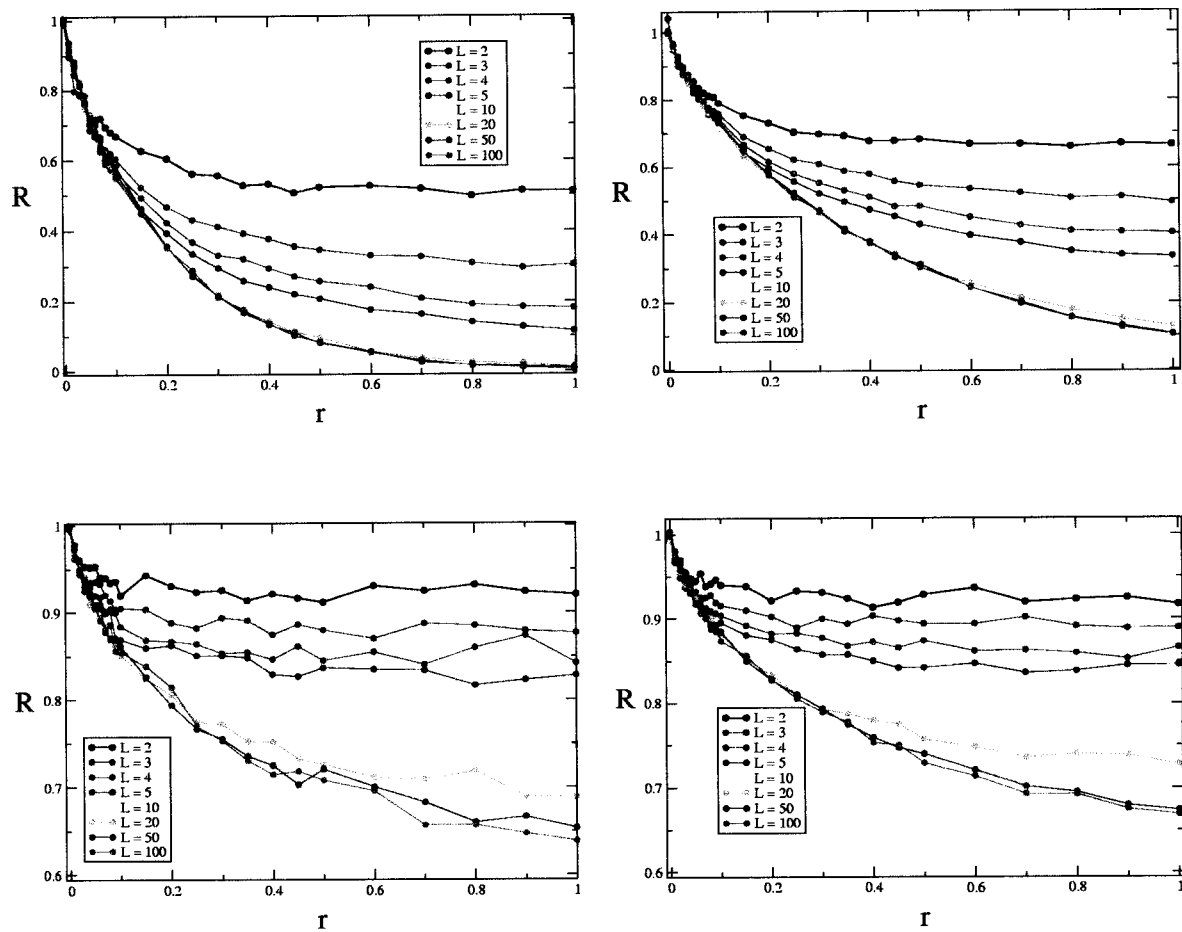


Figura 5.10: Dinâmica 1: velocidade da catraca em função do fator de recombinação r para L variando de 2 a 100. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$, com $U = \frac{0.5}{L}$ à esquerda e $U = \frac{1.0}{L}$ à direita. Médias feitas em $NPOP = 50$ populações de $N = 100$ indivíduos.

5 Simulações

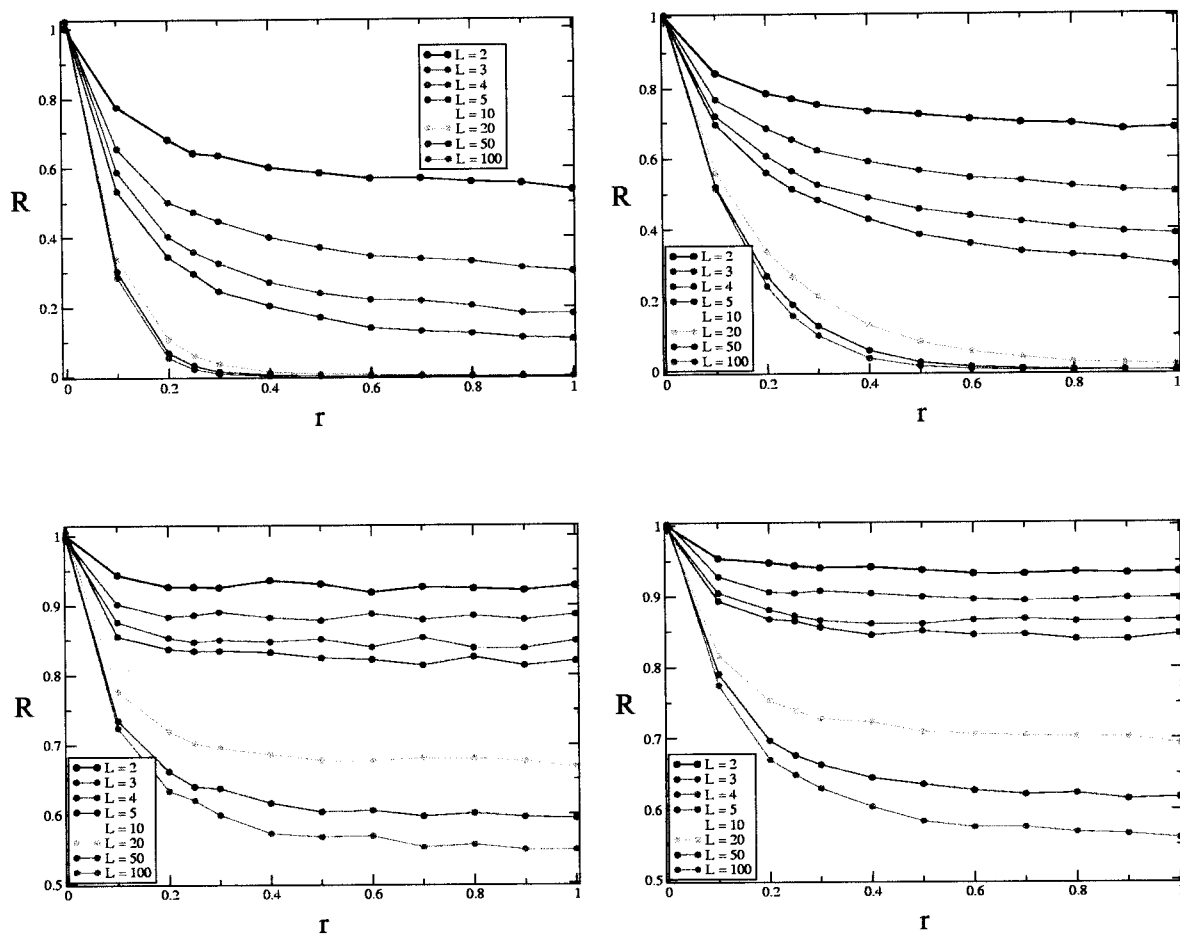


Figura 5.11: Dinâmica 2: velocidade da catraca em função do fator de recombinação r para L variando de 2 a 100. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$, com $U = \frac{0.5}{L}$ à esquerda e $U = \frac{1.0}{L}$ à direita. Médias feitas em $NPOP = 100$ populações de $N = 100$ indivíduos.

5 Simulações

tendência de saturação, como no caso anterior. Em termos qualitativos, a Dinâmica 2, na Figura 5.13, apresenta resultados semelhantes. Por limitações computacionais não pudemos avançar no tamanho L dos indivíduos, de modo que a avaliação correta do comportamento assintótico de R não pôde ser obtida. Ainda assim, a técnica de regressão linear foi utilizada para os maiores valores de L , ou seja, para os últimos pontos das simulações. Notamos (Figura 5.14) que a catraca tende a conter seu movimento apenas para alguns conjuntos de parâmetros, mantendo uma velocidade constante para outros. Na Figura 5.14, com $U = \frac{1.0}{L}$ e $s = 0.1$, a catraca pára somente quando a probabilidade de recombinação é elevada.

5 Simulações

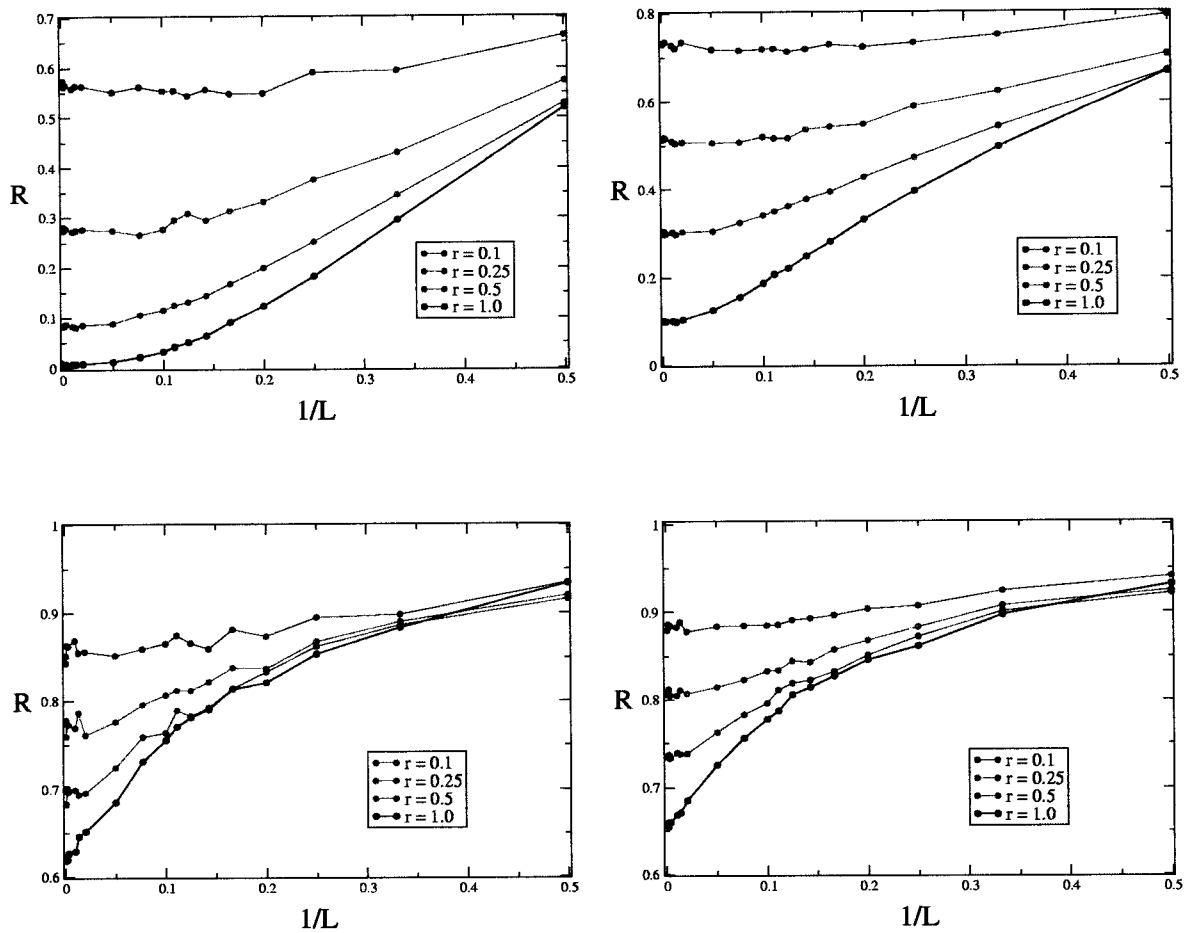


Figura 5.12: Dinâmica 1: velocidade da catraca em função do inverso de L para r variando de 0.1 a 1.0. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$, com $U = \frac{0.5}{L}$ à direita e $U = \frac{1.0}{L}$ à esquerda. Médias feitas em $NPOP = 100$ populações de $N = 100$ indivíduos.

5 Simulações

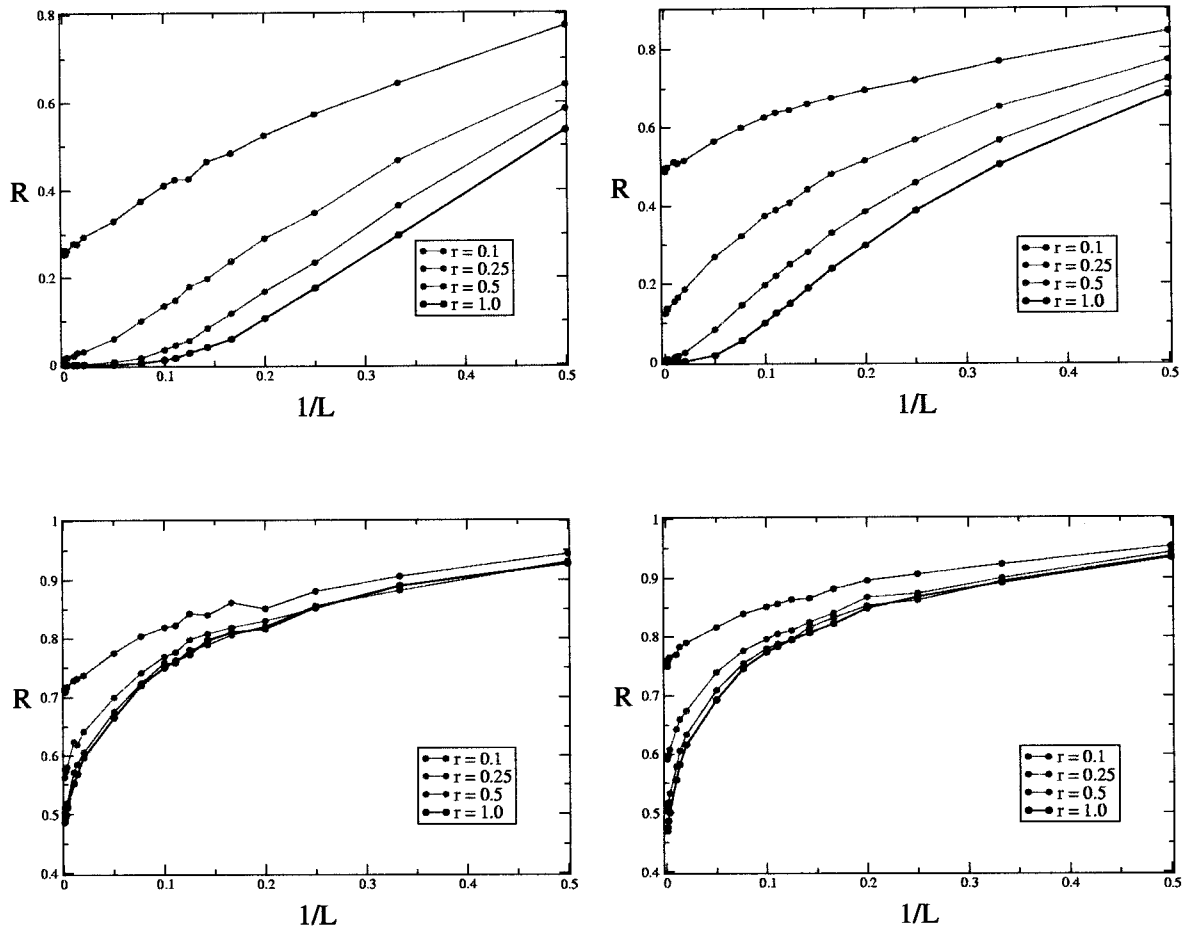


Figura 5.13: Dinâmica 2: velocidade da catraca em função do inverso de L para r variando de 0.1 a 1.0. Na parte superior, $s = 0.1$, e na parte inferior, $s = 0.01$, com $U = \frac{0.5}{L}$ à esquerda e $U = \frac{1.0}{L}$ à direita. Médias feitas em $NPOP = 100$ populações de $N = 100$ indivíduos.

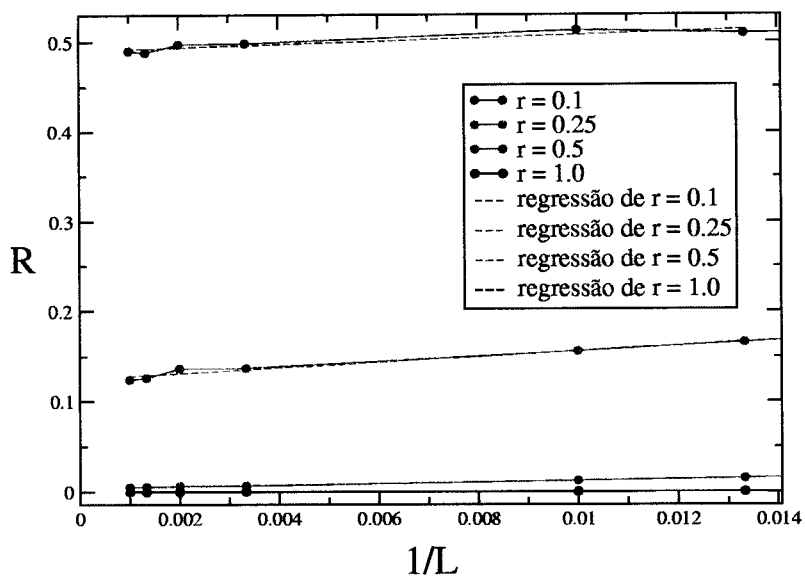


Figura 5.14: Dinâmica 2: Regressão linear da velocidade da catraca em função do inverso de L para r variando de 0.1 a 1.0. O parâmetro seletivo é $s = 0.1$ e $U = \frac{1.0}{L}$.

Capítulo 6

Considerações Finais

Vamos, agora, apresentar os principais resultados e conclusões obtidos nos capítulos anteriores. No Capítulo 2, apresentamos os modelos que estudamos nesta dissertação. No Capítulo 3, descrevemos o processo da catraca de Muller, enfatizando que ele só acontece em populações finitas. O limite determinístico, isto é, de população infinita, foi abordado no Capítulo 4, onde demos ênfase na obtenção das concentrações de equilíbrio de indivíduos com números determinados de mutações. Lá, utilizamos uma simplificação do modelo proposto no Capítulo 2, em que o número L de genes era variável; trabalhamos analiticamente somente com $L = 2$. Os resultados nos mostraram que as concentrações de equilíbrio dependem das condições iniciais da população, mas se $C_{jk}(t)$ se fatora em $C_j(t)C_k(t)$ em $t = 0$, a solução estacionária de $C_{jk}(t)$, C_{jk} , é dada pelo produto de C_j e C_k , ambos em equilíbrio. Numa extensão daquele capítulo, no Apêndice B, foi desenvolvida a solução para qualquer tempo de $C_j(t)$, que foi obtida de maneira muito elegante por um estudante de doutorado do grupo - Leonardo Paulo Maia - e que deu origem a uma publicação no periódico internacional *Journal of Mathematical Biology*.

Na seqüência da dissertação, Capítulo 5, foram expostos todos os resultados numéricos relevantes obtidos através de extensivas simulações. A ênfase foi dada ao comportamento da catraca sob situações diversificadas: aquelas em que variamos parâmetros como tamanho de indivíduo

6 Considerações Finais

(isto é, número de genes), de população, valor do coeficiente seletivo e das taxas de mutação e recombinação. Verificamos que a velocidade R da catraca é sensível a todos esses fatores, intensificada a sensibilidade quando de suas ações conjuntas. Como se esperava, maiores taxas de mutação aceleram o crescimento de R , enquanto maiores taxas de recombinação o retardam. Também pudemos notar que à medida que a população cresce ($N \rightarrow \infty$), R decresce, tendendo a zero rapidamente em alguns casos, mas mais lentamente em outros. Este resultado é satisfatório, haja vista a não existência do processo da catraca para populações infinitas. Infelizmente, não pudemos avançar muito em N por não possuímos técnicas de programação mais elaboradas. De fato, todas as nossas tentativas numéricas de obter expressões simples para o comportamento de R nos regimes assintóticos (N e/ou L grandes) falharam: na Figura 5.5, por exemplo, R parece diminuir exponencialmente com N para U pequeno ($U \leq 0.3$), mas essa conclusão deixa de ser válida para U grande. Parece-nos claro que, apesar das extremas simplificações frente à situação biológica real, o modelo estudado é ainda muito complexo (ao todo há cinco parâmetros de controle - N, L, s, u e r - todos relevantes) para ser descrito por relações simples, às quais os físicos estão habituados, mesmo nos regimes assintóticos.

Os resultados das duas dinâmicas (ou melhor, procedimentos de recombinação) estudadas concordam em todos os aspectos e são diferentes apenas quantitativamente, nos valores de $\langle k_{min} \rangle$ alcançados, em especial quando o parâmetro a ser variado é o tamanho L dos indivíduos. De modo geral, observamos que a catraca pode ser parada através da reprodução sexuada, evitando, assim, que populações sujeitas à ação de mutações deletérias sejam extintas. Em um experimento real, Chao [23] pôde verificar que o sexo realmente ajuda uma população a se livrar de mutações prejudiciais e, conseqüentemente, da catraca.

APÊNDICES

Apêndice A

A distribuição de Poisson

A probabilidade de que um evento caracterizado pela probabilidade p ocorra n vezes num total de N tentativas independentes é dada pela distribuição binomial,

$$W(n) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}. \quad (\text{A.1})$$

Considerando uma situação em que p seja pequeno ($p \ll 1$) e que, portanto, $W(n)$ seja apreciavelmente diferente de zero apenas para $n \ll N$, vamos mostrar que a distribuição binomial toma uma forma diferente e é chamada de distribuição de Poisson. Na natureza, um evento assim pode ser representado pela probabilidade de que aconteçam n mutações num indivíduo que possui N nucleotídeos, com $N \rightarrow \infty$ (N típico é da ordem de 10^9). Neste caso, p é a taxa de mutação por nucleotídeo, que vale em média 10^{-9} .

Podemos notar que com N grande e p pequeno, podemos simplificar a distribuição $W(n)$. Trabalhando-a por partes temos

$$\frac{N!}{(N-n)!} = \frac{N(N-1)(N-2)(N-3)\dots(N-n)!}{(N-n)!}.$$

Como $N \gg n$, $N-1 \sim N$, assim como todos os termos do numerador, isto é:

$$\frac{N!}{(N-n)!} = N^n. \quad (\text{A.2})$$

A A distribuição de Poisson

Se aplicarmos o logaritmo natural no termo $(1-p)^{N-n}$,

$$f(p) = \ln(1-p)^{N-n},$$

e o expandirmos em série de Taylor em torno da origem, $p = 0$ (pois $p \ll 1$), obtemos

$$\begin{aligned} f(p) &\approx f(0) + pf'(0) + \frac{p^2}{2!}f''(0) + \dots \\ f(p) &\approx 0 - p(N-n) - \frac{p^2}{2}(N-n) + \dots \end{aligned}$$

Desprezando termos que contenham potências de p iguais ou superiores a 2 (quando da multiplicação com N e n), encontramos

$$f(p) \approx -p(N-n).$$

Mas $N \gg n$, de modo que

$$f(p) \approx -pN. \tag{A.3}$$

Aplicando a exponencial em $f(p)$:

$$e^{f(p)} = e^{-pN} = e^{-\lambda},$$

onde λ é a bem conhecida média da distribuição binomial. Assim,

$$W_p(n) = \frac{N^n p^n e^{-pN}}{n!} = \frac{\lambda^n e^{-\lambda}}{n!}. \tag{A.4}$$

Vale lembrar que $W(n)$ é normalizada, pois

$$\sum_{n=0}^{\infty} W_p(n) = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda} e^{\lambda} = 1.$$

A média da distribuição de Poisson, por sua vez, é dada por

$$\langle n \rangle = \sum_{n=0}^{\infty} n \frac{\lambda^n e^{-\lambda}}{n!} = \sum_{n=1}^{\infty} n \frac{\lambda^n e^{-\lambda}}{n(n-1)!} = \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{(n-1)!}. \tag{A.5}$$

Multiplicando e dividindo por λ ,

$$\langle n \rangle = \lambda \sum_{n=1}^{\infty} \frac{\lambda^{n-1} e^{-\lambda}}{(n-1)!} = \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!}$$

A A distribuição de Poisson

e fazendo $m = n - 1$, então,

$$\langle m + 1 \rangle = \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = \lambda e^{-\lambda} e^{\lambda},$$

e

$$\langle n \rangle = \lambda. \quad (\text{A.6})$$

Da mesma maneira podemos calcular $\langle n^2 \rangle$,

$$\langle n^2 \rangle = \sum_{n=0}^{\infty} n^2 \frac{\lambda^n e^{-\lambda}}{n!} = \sum_{n=1}^{\infty} n^2 \frac{\lambda^n e^{-\lambda}}{n(n-1)!} = \sum_{n=1}^{\infty} n \frac{\lambda^n e^{-\lambda}}{(n-1)!}. \quad (\text{A.7})$$

Novamente multiplicando e dividindo a equação por λ ,

$$\langle n^2 \rangle = \lambda \sum_{n=1}^{\infty} n \frac{\lambda^{n-1} e^{-\lambda}}{(n-1)!},$$

e fazendo $m = n - 1$,

$$\begin{aligned} \langle (m+1)^2 \rangle &= \lambda \sum_{m=0}^{\infty} (m+1) \frac{\lambda^m e^{-\lambda}}{m!} \\ &= \lambda e^{-\lambda} \left\{ \sum_{m=0}^{\infty} m \frac{\lambda^m e^{-\lambda}}{m!} + \sum_{m=0}^{\infty} \frac{\lambda^m e^{-\lambda}}{m!} \right\} \\ &= \lambda e^{-\lambda} \{ \lambda e^{\lambda} + e^{\lambda} \} \\ &= \lambda(\lambda + 1), \end{aligned} \quad (\text{A.8})$$

obtemos

$$\langle n^2 \rangle = \lambda^2 + \lambda. \quad (\text{A.9})$$

A variância da distribuição é escrita como

$$\begin{aligned} \langle (\Delta n)^2 \rangle &= \langle n^2 \rangle - \langle n \rangle^2 \\ &= \lambda^2 + \lambda - \lambda^2, \end{aligned} \quad (\text{A.10})$$

de modo que

$$\langle (\Delta n)^2 \rangle = \lambda. \quad (\text{A.11})$$

A A distribuição de Poisson

Como gerar números aleatórios de acordo com a distribuição de Poisson

Os geradores comuns de números aleatórios seguem uma distribuição uniforme, isto é, geram números igualmente prováveis de serem obtidos dentro de um intervalo desejado, usualmente de 0 a 1. Sabemos que $W_p(n)$, para qualquer n , está no intervalo 0-1, e a soma em n de $W_p(n)$ é a unidade, como já foi demonstrado.

Vamos, agora, desenvolver um método para geração de números aleatórios com distribuição de probabilidade não mais uniforme, mas segundo uma Poisson. Para isto, daremos um exemplo concreto quando $\lambda = 1$:

- geramos um número aleatório da maneira usual (uniforme) entre zero e um. Vamos supor que ele valha 0.84;
- definimos a média da distribuição de Poisson que desejamos obter, neste caso, $\lambda = 1$;
- calculamos $P_0 = W_p(0) = 0.36$;
- analisamos: o número aleatório gerado se encontra no intervalo $0 - P_0$? Ou seja, entre 0 e 0.36? Não.
- calculamos $P_1 = W_p(0) + W_p(1) = 0.72$;
- analisamos: o número aleatório gerado se encontra no intervalo $P_0 - P_1$? Ou seja, entre 0.36 e 0.72? A resposta novamente é negativa.
- calculamos $P_2 = W_p(0) + W_p(1) + W_p(2) = 0.90$;
- analisamos: o número aleatório gerado se encontra no intervalo $P_1 - P_2$? Ou seja, entre 0.72 e 0.90? A resposta desta vez é positiva.
- dizemos, portanto, que o número de Poisson obtido vale 2, que é o índice do último P_n calculado;

A A distribuição de Poisson

- repetimos o processo quantas vezes desejarmos;

Seguindo este procedimento vamos encontrar problemas quando o número n de divisões do intervalo 0-1 ficar da ordem de 30, devido aos fatoriais envolvidos no cálculo de $W_p(n)$. É por isso que a partir deste valor crítico passamos a utilizar a fórmula de Stirling para os fatoriais:

$$\frac{1}{n!} \approx \left(\frac{e}{n}\right)^n \frac{1}{\sqrt{2n\pi}}, \quad (\text{A.12})$$

de maneira a escrevermos

$$W_p(n) = \left(\frac{\lambda}{n}\right)^n \frac{e^{n-\lambda}}{\sqrt{2n\pi}}. \quad (\text{A.13})$$

Quando n é muito grande já não é preciso dividir os intervalos de 0 a 1 porque as divisões são tão pequenas que a máquina não tem precisão suficiente para somar os termos de $W_p(n)$. Assim, definimos que a partir deste número limite de n , a soma de $W_p(n)$ não varia e vale 1.

Distribuição Simulada × Distribuição Teórica

Nos gráficos a seguir são comparados os valores teóricos e simulados da distribuição de Poisson. Três casos são mostrados: $\lambda = 0.1$ e $\lambda = 1$ (Fig A.1) e $\lambda = 10$ (Fig A.2). Para todos os gráficos foram simulados 10^6 pontos.

A A distribuição de Poisson

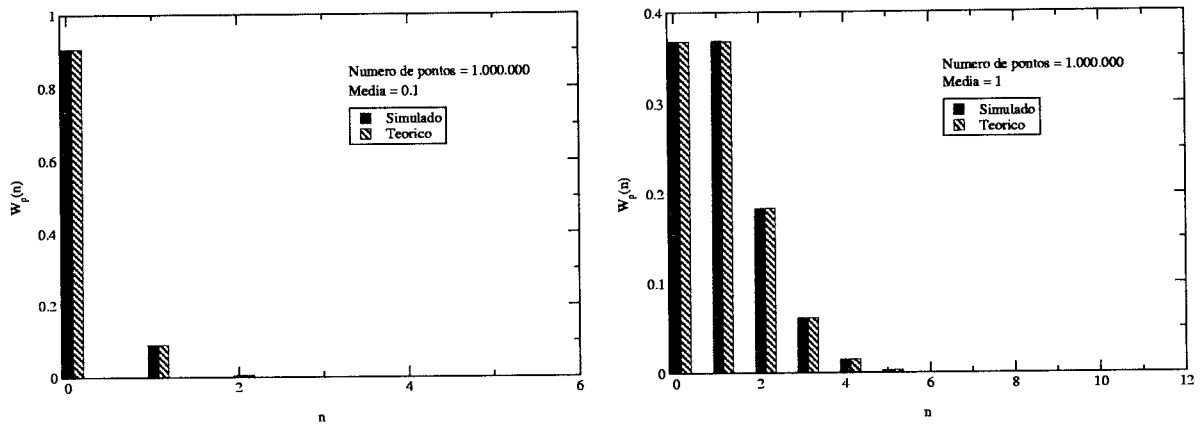


Figura A.1: Distribuição de Poisson com média 0.1 (à esquerda) e média 1 (à direita).

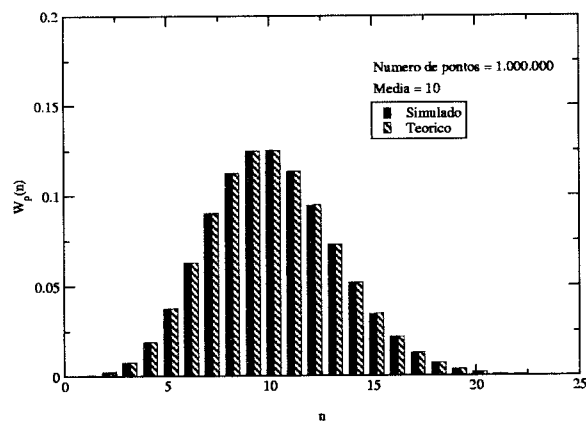


Figura A.2: Distribuição de Poisson com média 10.

Apêndice B

Solução analítica da dinâmica evolutiva

Uma população assexuada infinita evolui de acordo como o modelo de Wright-Fisher sujeita ao efeito de mutação e seleção [14]. Um indivíduo dessa população é representado por um gene de infinitos sítios, de modo que uma nova mutação nunca acontece num sítio já mutado. Esse processo garante a criação de novos indivíduos na população todas as vezes que uma mutação ocorre. A probabilidade que seja k o número de novas mutações que um indivíduo receba é dada por uma distribuição de Poisson,

$$M_k = e^{-U} \frac{U^k}{k!}, \quad (\text{B.1})$$

sendo U o número médio de novas mutações por indivíduo por geração. Consideramos que cada nova mutação reduz o valor da fitness pela mesma fração (modelo de fitness multiplicativa),

$$w_k = (1 - s)^k, \quad (\text{B.2})$$

em que $s \in [0, 1]$ mede o grau deletério das mutações.

Um indivíduo na geração $t + 1$ possui um número de mutações totais que é a soma das mutações herdadas do seu pai (pertencente à geração t) e das novas mutações que recebe, de

B Solução analítica da dinâmica evolutiva

acordo com a distribuição de Poisson. O modelo assume, além da não-superposição de gerações, que a probabilidade de um indivíduo gerar um filho na geração seguinte depende somente de w_k . Desta forma, a frequência de indivíduos na classe i - isto é, que possuem i mutações - na geração t é dada pela relação de recorrência [28]

$$C_i(t+1) = \frac{1}{w(t)} \sum_{j=0}^i M_{i-j} w_j C_j(t), \quad (\text{B.3})$$

sendo $w(t) = \sum_{j=0}^{\infty} w_j C_j(t)$ a fitness média da população na geração t .

O balanço entre seleção e mutação gera uma distribuição de equilíbrio das mutações na população que é descrita pela solução estacionária da eq(B.3), quando $C_j(t) = C_j$. Como foi demonstrado no capítulo 4, a solução estacionária para o problema, ao se considerar que somente as frequências das classes com $j \geq i$ são diferentes de zero, é dada pela eq(4.22),

$$C_j = e^{-\theta} \frac{\theta^{j-i}}{(j-i)!} \quad j \geq i, \quad (\text{B.4})$$

onde vale lembrar que, aqui, $\theta = \frac{U}{s}$. Este é um resultado conhecido [28], cuja dedução é feita supondo-se $C_j(t+1) = C_j(t) = C_j$.

Nesse apêndice mostraremos uma solução para a eq(B.3) que é válida para todos os tempos e frequências iniciais $C_i(t=0)$. Para isso, vamos definir

$$G(z, t) = \sum_{j=0}^{\infty} C_j(t) z^j, \quad (\text{B.5})$$

de modo que

$$w(t) = \sum_{j=0}^{\infty} C_j(t) (1-s)^j = G(1-s, t). \quad (\text{B.6})$$

Aqui, o índice varia de 0 a ∞ pois podem existir classes de indivíduos com qualquer número j de mutações, já que o modelo é de infinitos alelos. A relação de recorrência para $C_j(t+1)$ é

$$\begin{aligned} C_j(t+1) &= \sum_{m=0}^j C_m(t) M_{j-m} \frac{w_m}{w(t)} \\ C_j(t+1) &= \sum_{m=0}^j C_m(t) \frac{e^{-U} U^{j-m} (1-s)^m}{(j-m)! w(t)}. \end{aligned} \quad (\text{B.7})$$

B Solução analítica da dinâmica evolutiva

Também temos que

$$G(z, t + 1) = \sum_{j=0}^{\infty} C_j(t + 1) z^j = \sum_{j=0}^{\infty} \sum_{m=0}^j \frac{C_m(t)}{w(t)} \frac{e^{-U} U^{j-m}}{(j-m)!} (1-s)^m z^j, \quad (\text{B.8})$$

que, multiplicada por $z^m z^{-m}$, fornece

$$G(z, t + 1) = \sum_{j=0}^{\infty} \sum_{m=0}^j \frac{C_m(t)}{w(t)} \frac{e^{-U} U^{j-m}}{(j-m)!} (1-s)^m z^{j-m} z^m. \quad (\text{B.9})$$

Trabalhando com as somas e fazendo $l = j - m$, podemos escrever

$$\begin{aligned} G(z, t + 1) &= \frac{e^{-U}}{w(t)} \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} C_m(t) \frac{(Uz)^l}{l!} [z(1-s)]^m \\ &= \frac{e^{-U}}{w(t)} \sum_{m=0}^{\infty} C_m(t) [z(1-s)]^m e^{Uz} \\ &= \frac{e^{-U}}{w(t)} G[z(1-s), t] e^{Uz} \\ &= \frac{e^{U(z-1)}}{w(t)} G[z(1-s), t] \\ &= e^{U(z-1)} \frac{G[z(1-s), t]}{G(1-s, t)}. \end{aligned} \quad (\text{B.10})$$

Aplicando a eq(B.10) nela mesma obtemos

$$G(z, t + 1) = e^{U(z-1)(2-s)} \frac{G[z(1-s)^2, t-1]}{G[(1-s)^2, t-1]}. \quad (\text{B.11})$$

Uma fórmula geral para a eq(B.11) poderia ser escrita da seguinte maneira:

$$G(z, t + 1) = e^{U(z-1)\theta_m} \frac{G[z(1-s)^m, t-m+1]}{G[(1-s)^m, t-m+1]}, \quad (\text{B.12})$$

sendo $\theta_1 = 1$ e $\theta_2 = 2 - s$. Aplicando a eq(B.10) no segundo membro de (B.12),

$$G(z, t + 1) = e^{U(z-1)[\theta_m + (1-s)^m]} \frac{G[z(1-s)^{m+1}, t-m]}{G[(1-s)^{m+1}, t-m]}, \quad (\text{B.13})$$

de onde tiramos que $\theta_{m+1} = \theta_m + (1-s)^m$. Assim,

$$\begin{aligned} \theta_m &= \theta_{m-2} + (1-s)^{m-2} + (1-s)^{m-1} \\ &= \theta_{m-3} + (1-s)^{m-3} + (1-s)^{m-2} + (1-s)^{m-1}. \end{aligned} \quad (\text{B.14})$$

B Solução analítica da dinâmica evolutiva

Se assumimos $\theta_0 = 0$,

$$\begin{aligned}\theta_m &= \theta_0 + (1-s)^0 + (1-s)^1 + (1-s)^2 + (1-s)^3 + \dots + (1-s)^{m-1} \\ &= \sum_{j=0}^{m-1} (1-s)^j,\end{aligned}\tag{B.15}$$

que é uma progressão geométrica finita de razão $(1-s)$, cuja soma vale

$$\theta_m = \frac{1 - (1-s)^m}{s}.\tag{B.16}$$

Voltando esse resultado na eq(B.13)

$$G(z, t+1) = e^{U(z-1)\theta_{m+1}} \frac{G[z(1-s)^{m+1}, t-m]}{G[(1-s)^{m+1}, t-m]}\tag{B.17}$$

e fazendo $t = m$, temos

$$G(z, t+1) = e^{U(z-1)\theta_{t+1}} \frac{G[z(1-s)^{t+1}, 0]}{G[(1-s)^{t+1}, 0]},\tag{B.18}$$

que é o mesmo que escrever

$$G(z, t) = e^{U(z-1)\theta_t} \frac{G[z(1-s)^t, 0]}{G[(1-s)^t, 0]}.\tag{B.19}$$

Da definição da geratriz, eq(B.5),

$$G(z, t) = e^{U(z-1)\theta_t} \frac{\sum_{j=0}^{\infty} C_j(0)[z(1-s)^t]^j}{\sum_{j=0}^{\infty} C_j(0)(1-s)^{jt}}.\tag{B.20}$$

Expandindo o lado direito dessa equação em série de Taylor obtemos, finalmente,

$$C_j(t) = \frac{e^{-U\theta_t}}{\sum_{j=0}^{\infty} C_j(0)(1-s)^{jt}} \sum_{i=0}^j \frac{(U\theta_t)^{j-i}}{(j-i)!} C_i(0)(1-s)^{it}.\tag{B.21}$$

Este mesmo resultado poderia ter sido obtido através da relação

$$C_j(t) = \frac{1}{j!} \frac{\partial^j}{\partial z^j} [G(z, t)]_{z=0},\tag{B.22}$$

e é o que vamos mostrar agora. A dificuldade do método está no cálculo das j -ésimas derivadas de um produto. Utilizando a igualdade a seguir,

$$(f \cdot g)^{(k)} = \sum_{i=0}^k \frac{k!}{i!(k-i)!} f^{(k-i)} g^{(i)}\tag{B.23}$$

B Solução analítica da dinâmica evolutiva

na eq(B.20),

$$\frac{\partial^k G(z, t)}{\partial z^k} = \frac{e^{-U\theta t}}{\sum_{j=0}^{\infty} C_j(0)(1-s)^{jt}} \sum_{i=0}^k \frac{k!}{i!(k-i)!} \left[e^{Uz\theta t} \right]^{(k-i)} \left[\sum_{j=0}^{\infty} C_j(0)(1-s)^{jt} z^j \right]^{(i)}. \quad (\text{B.24})$$

Como a derivada da soma é a soma das derivadas,

$$\frac{\partial^k G(z, t)}{\partial z^k} = \frac{e^{-U\theta t}}{\sum_{j=0}^{\infty} C_j(0)(1-s)^{jt}} \sum_{i=0}^k \frac{k!}{i!(k-i)!} (U\theta t e^{Uz\theta t})^{k-i} \sum_{j=0}^{\infty} C_j(0)(1-s)^{jt} [z^j]^{(i)}, \quad (\text{B.25})$$

que se resume a

$$\frac{\partial^k G(z, t)}{\partial z^k} = \frac{e^{-U\theta t}}{\sum_{j=0}^{\infty} C_j(0)(1-s)^{jt}} \sum_{i=0}^k \frac{k!}{i!(k-i)!} (U\theta t e^{Uz\theta t})^{k-i} \sum_{j=0}^{\infty} C_j(0)(1-s)^{jt} z^{j-i} \frac{j!}{(j-i)!} \quad (\text{B.26})$$

onde utilizamos o fato de que

$$[z^j]^{(i)} = z^{j-i} \frac{j!}{(j-i)!}. \quad (\text{B.27})$$

Tomando $z = 0$, a última soma do membro direito da eq(B.26) só é não nula quando $j = i$. Essa condição faz com que a soma desapareça e dê lugar a um único termo, aquele para o qual $j = i$,

$$\begin{aligned} \frac{\partial^k G(z, t)}{\partial z^k} &= \frac{e^{-U\theta t}}{\sum_{j=0}^{\infty} C_j(0)(1-s)^{jt}} \sum_{i=0}^k \frac{k!}{i!(k-i)!} (U\theta t)^{k-i} C_i(0)(1-s)^{it} i! \\ &= \frac{e^{-U\theta t}}{\sum_{j=0}^{\infty} C_j(0)(1-s)^{jt}} \sum_{i=0}^k k! \frac{(U\theta t)^{k-i}}{(k-i)!} C_i(0)(1-s)^{it}. \end{aligned} \quad (\text{B.28})$$

Voltando esse resultado na eq(B.22) e trocando-se os índices obtemos

$$C_j(t) = \frac{e^{-U\theta t}}{\sum_{j=0}^{\infty} C_j(0)(1-s)^{jt}} \sum_{i=0}^j \frac{(U\theta t)^{k-i}}{(k-i)!} C_i(0)(1-s)^{it}. \quad (\text{B.29})$$

Por fim, a fitness média da população na geração t é dada por $G(1-s, t)$ na eq(B.19),

$$w(t) = \frac{e^{-U\theta t s}}{\sum_{j=0}^{\infty} C_j(0)(1-s)^{jt}} \sum_{j=0}^{\infty} C_j(0)(1-s)^{j(t+1)}. \quad (\text{B.30})$$

Apêndice C

Probabilidade de recombinantes

Um casal $A_1A_2\dots A_L$ e $B_1B_2\dots B_L$, em que os índices inferiores se referem aos genes cujos alelos estão associados, é escolhido da população na geração t para gerar um indivíduo que pertencerá à geração $t + 1$. O 1º gene filial é escolhido com 50% de chance de vir do pai ou da mãe. O mesmo acontece para o restante dos genes, até que todo o genótipo do indivíduo seja formado. Nessa formulação, atribuímos a r a probabilidade uniforme de haver ou não recombinação. Assim, a probabilidade que o filho seja idêntico a um dos pais é dada por

$$P[A_1A_2\dots A_L] = P[B_1B_2\dots B_L] = \frac{1-r}{2}. \quad (\text{C.1})$$

A probabilidade que o filho herde uma quantidade i de genes maternos e uma quantidade $L - i$ de genes paternos é escrita como

$$P[iA's, (L-i)B's] = r \left(\frac{1}{2}\right)^L \quad (\text{C.2})$$

e é independente do valor de i . Duas seqüências especiais, aquelas que possuem só $A's$ ou só $B's$ até o gene $L - 1$ têm probabilidades distintas de acontecer, dadas por

$$P[A_1A_2\dots A_{L-1}B_L] = P[B_1B_2\dots B_{L-1}A_L] = r \left(\frac{1}{2}\right)^{L-1}. \quad (\text{C.3})$$

Isso acontece porque se $r = 1$, isto é, se houver recombinação, sabemos que o genótipo filial não pode ser idêntico ao parental. Assim, devemos impor a condição de que se os $L - 1$ genes filiais

C Probabilidade de recombinantes

forem provenientes de somente um indivíduo da geração parental, o pai, por exemplo, seu gene L deve provir da mãe (e vice-versa), sem chance de escolha. Dessa forma, é como se tivesse sido retirado um termo $\frac{1}{2}$ da equação (C.2), devido à condição imposta. No Capítulo 2, Figura 2.2, $L = 4$, de maneira a resumirmos esses resultados em

$$\begin{aligned}P[A_1A_2A_3A_4] &= P[AAAA] = \frac{1-r}{2} \\P[B_1B_2B_3B_4] &= P[BBBB] = \frac{1-r}{2} \\P[A_1A_2A_3B_4] &= P[AAAB] = r \left(\frac{1}{2}\right)^3 \\P[B_1B_2B_3A_4] &= P[BBBA] = r \left(\frac{1}{2}\right)^3 \\P[AABB] &= P[BBAA] = P[ABAB] = P[BABA] = r \left(\frac{1}{2}\right)^4, \quad (C.4)\end{aligned}$$

Como foi mencionado no Capítulo 2, esse é apenas um exemplo ilustrativo já que em nossas simulações utilizamos o modelo de infinitos alelos, e não de dois, A e B .

Bibliografia

- [1] www.ib.usp.br/textos
- [2] L. E. Mettler e T. G. Gregg , *Genética de Populações e Evolução*, Editora da Universidade de São Paulo, São Paulo,1973
- [3] www.dreyfus.ib.usp.br/bio212/genpop
- [4] M. Kimura, *On the probability of fixation of mutant genes in a population*, *Genetics* **47** 713-719 (1962)
- [5] H. J. Muller, *The relation of recombination to mutational advantage*, *Mutat. Res.* **1 2** (1964)
- [6] B. Drossel, *Biological evolution and statistical physics*, *Adv. Phys.* **50** (2) 209-295 (Mar 2001). Disponível para *download* em <http://br.arxiv.org/abs/cond-mat/0101409>.
- [7] E. Baake and W. Gabriel, *Biological evolution through mutation, selection and drift: an introductory review* *Ann. Rev. Comp. Phys.* (D. Stauffer, ed), World Scientific, vol 9, 203-264 (2000). Disponível para *download* em <http://br.arxiv.org/abs/cond-mat/9907372>.
- [8] G. Bell, *Recombination and the immortality of the germ line*, *J. Evol. Biol.* **1** 67-82 (1988)
- [9] P. G. Higgs, *Error thresholds and stationary mutant distribution in multi-locus diploid genetic models*, *Gent. Res.* **63** 63-78 (1994)

BIBLIOGRAFIA

- [10] P. G. Higgs and G. Woodcock, *The accumulation of mutations in asexual populations and the structure of genealogical trees in the presence of selection*, *Journal of Mathematical Biology* **33** 677-702 (1995)
- [11] G. Bell, *Sex and death in protozoa. The history of an Obsession*. Cambridge University Press, New York, 1989.
- [12] L. Chao, *Evolution of sex and the molecular clock in RNA viruses*, *Gene* **205** 301-308 (1997)
- [13] W. J. Ewens, *Mathematical Population Genetics*, Springer-Verlag, Berlin Heidelberg New York (1979)
- [14] D. L. Hartl and A. G. Clark, *Principles of Population Genetics*. (Sinauer Associates, Sunderland, 1989)
- [15] M. Kimura, *Theoretical foundation of population genetics at the molecular level*, *Theoret. Pop. Biol.* **2** 174 (1971)
- [16] D. E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley, Reading, 1989.
- [17] R. S. Howard and C. M. Lively , *Parasitism, mutation accumulation and the maintenance of sex*. *Nature* **367** 554-557 (1994)
- [18] R. S. Howard and C. M. Lively , *The Ratchet and the Red Queen: the maintenance of sex in parasites*. *J. Evol. Biol.* **15** 648-656 (2002)
- [19] J. F. Crow and M. Kimura, *An introduction to population genetics theory*. (Harper and Row, New York, 1970)
- [20] P. D. Keightley and A. Eyre-Walker, *Deleterious mutations and the evolution of sex*, *Science* **290** 331-333 (2000)

BIBLIOGRAFIA

- [21] J. M. Smith, *The evolution of sex*, Cambridge University Press, New York, 1978.
- [22] D. Charlesworth, M. T. Morgan and B. Charlesworth, *Mutation accumulation in finite populations*, *Journal of Heredity* **84** 321-325 (1993)
- [23] L. Chao, *Fitness of RNA virus decreased by Muller's ratchet*, *Nature* **348** 454-455 (29 Nov. 1990)
- [24] M. Lynch and W. Gabriel, *Mutation load and the survival of small populations*, *Evolution* **44** 1725-1737 (1990)
- [25] A. T. Bernardes, *Mutational meltdown in large sexual populations* *J. Phys. I* **5**(11) 1501-1515 (Nov. 1995)
- [26] P. Pamilo, M. Nei and W. H. Li, *Accumulation of mutations in sexual and asexual populations*, *Genet Res* **49**(2) 135-146 (Apr. 1987)
- [27] W. Gabriel, M. Lynch and R. Bürger, *Muller's ratchet and mutational meltdowns*. *Evolution*, **47**(6) 1744-1757 (1993)
- [28] J. Haigh, *The accumulation of deleterious genes in a population - Muller's ratchet*, *Theoret. Pop. Biol.* **14** 251-267 (1978)
- [29] L. P. Maia, D. F. Botelho and J. F. Fontanari, *Analytical solution of the evolution dynamics on a multiplicative-fitness landscape*, Aceito para publicação em *J. Math. Biol.* (Digital Object Identifier (DOI): 10.1007/s00285-003-0208-8)