

Introdução à Genética Matemática

Telles Timóteo da Silva¹ Thamara Carvalho Coutinho²

Campus Alto Paraopeba, Universidade Federal de São João del-Rei

¹ timoteo@ufsj.edu.br

² thamaraccoutinho@gmail.com

RESUMO

A genética de populações estuda a evolução do conjunto de genes de uma população de indivíduos, quando sujeitos às forças que tendem a modificar o conteúdo genético desta população. As forças principais habitualmente consideradas são a mutação, seleção natural e a deriva genética. Os modelos matemáticos são elaborados para tentar explicar e prever a forma como se dá esta evolução. Neste minicurso, veremos os conceitos básicos em genética populacional e os modelos de Hardy-Weinberg, de Wright-Fisher e de Moran. Vamos, assim, introduzir as principais bases matemáticas para a modelagem de problemas em genética de populações abordando os modelos clássicos e seus principais resultados.

Palavras-chave: Modelagem, Genética Populacional, Modelo de Hardy-Weinberg, Modelo de Wright-Fisher, Modelo de Moran.

Prefácio

Muitos seres deixaram de dar o primeiro passo, e, assim, não puderam dar o segundo...

Farid Ud-din Attar, *A Conferência dos Pássaros*

A genética de populações estuda a evolução do conjunto de genes de uma população de indivíduos, quando sujeitos às forças que tendem a modificar o conteúdo genético desta população [2, 10]. Os modelos matemáticos são elaborados para tentar explicar e prever a forma como se dá esta evolução [6, 19, 23, 27, 28]. As forças principais habitualmente consideradas que atuam no mesmo locus genético de todos os indivíduos são mutação, seleção natural e deriva genética, enquanto que outros fatores como a recombinação, a conversão e a inversão gênica necessitam de modelos que representem vários loci [5]. Já para modelar o efeito de migração é necessário que se considere diversas populações, onde uma eventualmente cede ou recebe indivíduos de populações vizinhas.

Pode-se dividir os modelos em dois grupos: determinísticos e estocásticos. Modelos determinísticos pressupõem que a população é formada por um número muito grande de indivíduos (em geral, infinita), mas não leva em conta possíveis flutuações no ambiente [15, 17]. Por sua vez, os modelos estocásticos não precisam fazer suposição sobre o tamanho da população, e além disso, trabalham naturalmente com flutuações aleatórias do ambiente, por isso tendem a ter mais aceitação nas aplicações. Também os modelos podem ser contínuos ou discretos no tempo e no espaço dos tipos de genes. Modelos estocásticos contínuos no tempo têm sua origem, em geral, como aproximações difusivas para modelos discretos [9, 20], e as formulações matemáticas que deles tratam requerem um conhecimento de técnicas sobre operadores diferenciais parciais [11, 12, 24], e processos estocásticos [3, 4, 7, 8, 13, 16].

O estudo de modelos em genética populacional se faz cada vez mais importante. Verifica-se que mais e mais dados experimentais têm sido postos à disposição dos pesquisadores [27, 28]. Dessa forma, é necessário se ter um ferramental matemático mais desenvolvido a fim de lidar com os dados experimentais.

A presente apostila foi composta a partir de importantes referências na área de Genética Matemática [1, 2, 5, 9, 14, 15, 22, 29]. O objetivo é tratar da modelagem matemática a partir de idéias simples, ao mesmo tempo em que se apresenta resultados interessantes do ponto de vista aplicado. O nível de conhecimento em matemática necessário para acompanhar os resultados é equivalente ao de um semestre de cálculo I, e de um semestre de estatística e probabilidade. O conteúdo de genética apresentado no texto pretende ser suficiente para a compreensão dos modelos apresentados. Ao longo do texto são propostos exercícios que motivem o estudante a pensar nos conceitos apresentados.

Ouro Branco, 19 de setembro de 2010.

Telles Timóteo da Silva
Thamara Carvalho Coutinho

Capítulo 1

Fundamentos de Genética

O meu bom senso não me diz o que é, mas deixa claro que há algo que precisa ser sabido.

Paulo Freire, *Pedagogia da Autonomia*

1.1 Ácidos nucleicos

A capacidade de armazenar e transmitir informação genética de uma geração para a seguinte é condição fundamental para a vida. As moléculas responsáveis por isso são conhecidas como **ácidos nucleicos**. A base dos ácidos nucleicos são os **nucleotídeos**. Esses possuem três componentes: uma base nitrogenada, uma pentose e um grupo fosfato [21].

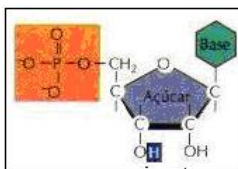


Figura 1.1: Nucleotídeo

Tanto o RNA quanto o DNA contêm duas bases púricas principais, a **adenina** (A) e a **guanina** (G) e duas pirimidinas. Em ambas uma das pirimidinas é a **citossina** (C), mas a outra não é a mesma: ela é **timina** (T)

para o DNA e **uracila** (U) para o RNA. Os ácidos nucleicos possuem duas espécies de pentoses. As unidades estruturais do DNA são chamadas **desoxirribonucleotídeos** e as unidades do RNA são os **ribonucleotídeos**. Os nucleotídeos sucessivos são unidos covalentemente por meio de pontes de grupos de fosfato, onde o grupo 5'- fosfato de uma unidade nucleotídica está unido ao grupo 3'-hidroxila do nucleotídeo seguinte, criando a ligação fosfodiéster [21].

O pareamento entre adenina e timina, e entre guanina e citosina no DNA, resulta numa orientação de complementariedade entre a sequência de bases nas duas cadeias entrelaçadas e fornece ao DNA seu caráter autocodificador. Por exemplo, se a sequência 5'-ATGTC-3' ocorre em uma cadeia, a cadeia oposta deverá apresentar a sequência complementar 3'-TACAG-5'. O RNA não é material genético e não precisa servir como molde para sua própria replicação. Em todas as funções já citadas do RNA ele é copiado como fita simples, produzida a partir de apenas uma das fitas do DNA molde, e não existe uma fita complementar a ela [26].

1.2 Termos usuais para compreensão da genética

1.2.1 Cromossomo

O DNA se encontra compactado dentro da célula numa estrutura chamada **cromossomo**. O DNA cromossômico é extremamente estável, permitindo que a informação codificada pelo DNA seja transmitida com segurança. As células procarióticas têm um cromossomo circular único, enquanto as células eucarióticas têm múltiplos cromossomos lineares

Cada célula mantém um número característico de cromossomos. A maioria das células eucarióticas é **diplóide**, ou seja, contém duas cópias de cada cromossomo. As duas cópias de um determinado cromossomo são chamadas de homólogos, sendo cada uma derivada de um progenitor. No entanto, um subconjunto de células eucarióticas pode ser **haplóide** ou **poliplóide**. As células haplóides contêm uma única cópia de cada cromossomo e estão envolvidas na reprodução celular (espermatozóides e óvulos). As células poliplóides possuem mais de duas cópias de cada cromossomo [26].

A espécie *Homo sapiens* possui 22 cromossomos mais os cromossomos haplóides X e Y, sendo que cada célula possui duas cópias de cada cromossomo, ou seja, o *Homo sapiens* é diplóide.

Termos usuais para compreensão da genética

1.2.2 Gene, locus e alelo

Gene corresponde a um pedaço de DNA (ver Seção 1.3.2) o qual contém uma informação genética. A localização de um gene é chamado **locus** e a forma alternativa do gene no locus é chamado **alelo** [15].

1.2.3 Genótipo e Fenótipo

A composição genética de um indivíduo é chamada **genótipo**, enquanto a sua aparência ou estrutura física é chamado **fenótipo**. Indivíduos com fenótipos idênticos podem apresentar genótipos diferentes, dessa forma, para determinar o fenótipo de um indivíduo é necessário realizar cruzamentos genéticos por várias gerações [26].

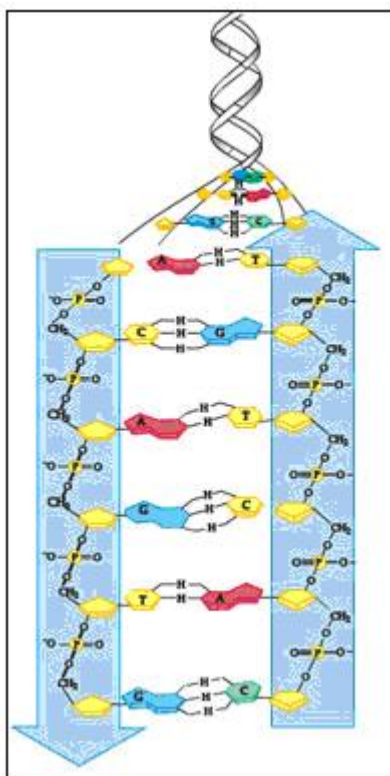


Figura 1.2: Estrutura tridimensional do DNA

Um par de genes no qual ambos os genes, materno e paterno, são idênticos são chamados **homozigotos**, enquanto pares com genes diferentes são chamados **heterozigotos**.

1.2.4 Dominância e codominância

Quando um alelo é expresso preferencialmente sobre outro ele é chamado de **dominante** e o outro de **recessivo**. Se um gene é heterozigoto Aa , onde A é dominante sobre a , o fenótipo é do tipo AA . Se ambos os alelos são expresso em heterozigose, produzindo um terceiro fenótipo, eles estão em estado de **codominância**.

1.2.5 Polimorfismo

As diferentes formas na qual um gene é encontrado na população é dada pelos alelos. A coexistência de dois ou mais alelos para determinar um único gene numa população é chamada **polimorfismo** [25].

É de interesse de muitos cientistas estudos de padrões polimórficos que foram distribuídos ao longo das culturas humanas, refletidos nas divisões dos continentes e nas ondas de migrações. Hoje é conhecido mais de 10 milhões de nucleotídeos polimórficos no genoma humano.

1.3 O dogma central

Em 1956, Francis Crick se referiu ao processo de transmissão da informação genética como sendo o dogma central [26].

As setas na figura 3 indicam as direções propostas para a transmissão da informação genética. A seta circundando o DNA significa que o DNA é o molde para sua própria replicação. A seta entre o DNA e o RNA indica que o DNA é molde para a síntese de RNA (transcrição). Da mesma forma a síntese de proteínas (tradução) é coordenada por um molde de RNA.

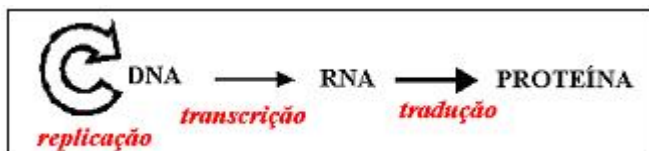


Figura 1.3: Dogma Central

O dogma central

1.3.1 Replicação do DNA

A estrutura complementar das bases na molécula do DNA é essencial para a sua replicação, na qual cada fita serve como molde para a formação de uma fita filha complementar. Contudo a replicação da molécula de DNA mais simples é um processo complexo, de múltiplas etapas, que envolvem muitas enzimas.

A síntese de DNA depende da presença de dois tipos de substratos: os quatro desoxinucleosídeos trifosfatados, cada um correspondente a uma base: dATP, dGTP, dTTP e dCTP; e uma estrutura molde de DNA. O molde de DNA determina a seqüência de nucleotídeos incorporados. O iniciador atua como substrato para a adição de desoxinucleotídeos, os quais são sucessivamente adicionados ao grupo 3'OH na extremidade 3' do iniciador.

A síntese de DNA é catalizada pela enzima DNA-polimerase, a qual se liga ao sítio catalítico. Essa enzima atua de maneira processiva: uma vez ligada a um substrato é capaz de adicionar muitos nucleotídeos. Uma outra enzima conhecida por exonuclease faz a revisão da leitura, atuando como removedora de nucleotídeos adicionados incorretamente. Ambas as fitas do molde de DNA são duplicadas simultaneamente em uma estrutura chamada forquilha de replicação. Como as duas fitas são antiparalelas, apenas uma das fitas do molde pode ser replicada de maneira contínua, a que cresce no sentido 5' para 3'. A outra fita de DNA deve ser sintetizada primeiramente como uma série de pequenos fragmentos recém sintetizados conhecidos como fragmentos de Okazaki. Cada fita de DNA é iniciada com um iniciador de RNA(primer), que é sintetizado pela enzima primase. Esses iniciadores devem ser removidos para finalizar o processo de replicação. Após a substituição dos RNAs iniciadores por DNA, todos os fragmentos de Okazaki são unidos covalentemente produzindo uma fita contínua de DNA.

1.3.2 Transcrição do RNA e tradução de proteínas

A transcrição é o processo de síntese de RNA a partir do DNA. Ela é quimicamente e enzimaticamente semelhante à replicação do DNA. Os mecanismos que diferem os dois processos são os seguintes:

- i) A enzima que sintetiza a nova fita, RNA polimerase, adiciona ribonucleotídeos e não necessita de um iniciador (*primer*). Ela precisa apenas de alguns fatores de iniciação que asseguram que a enzima inicie a transcrição em sítios apropriados do DNA chamados promotores.
- ii) O RNA produzido não permanece ligado pelas bases ao DNA molde, a enzima libera a cadeia em crescimento, o que é fundamental para a tradução em proteínas.

iii) A transcrição é menos precisa que a replicação. Isso se deve à ausência do mecanismo geral de revisão da leitura.

Um ciclo de transcrição envolve três fases: iniciação, alongamento e terminação. Durante a iniciação a RNA polimerase se liga ao promotor, formando um complexo fechado. A seguir, o DNA em torno do sítio de iniciação é desenrolado e acontece a síntese de uma série de RNAs curtos. Com isso o promotor escapa e a enzima inicia a fase de alongamento na qual ela: abre o DNA jusante e o recompõe a montante, adiciona ribonucleotídeos à extremidade 3' (cerca de 8 ou 9), remove o RNA recém formado e corrige o transcrito, verificando os nucleotídeos incorretamente inseridos (de maneira menos eficiente do que na replicação).

A sequência codificante de um gene é a série de códons, compostos por três nucleotídeos, que ditarão a sequência linear de aminoácidos. Porém no DNA genômico existem algumas regiões que não são codificantes, chamadas íntrons, enquanto as regiões codificantes são os éxons. Quando um gene com íntron é transcrito, o RNA inicial que contém esses íntrons passa por um processamento aonde eles são removidos para produzir o RNA maduro.

A maquinaria para a síntese proteica é composta por quatro componentes principais: RNAm, RNAt, aminoácil-RNAt sintetase e o ribossomo (complexo com várias subunidades que cataliza as ligações peptídicas). O RNAm contém a sequência codificante (códon); os elementos de reconhecimento para iniciação, códon AUG; e o de terminação, repetição de resíduos com adenina (GAA, AGA). As aminoácil-RNAt sintetase ligam os aminoácidos aos RNAt. O RNAt apresentam o anticódon que é complementar aos códons do RNAm, reconhecido por pareamento de bases. Esses anticódons são os **aminoácidos** que formam ligações peptídicas entre si. Todos esses processos acontecem no ribossomo. Na finalização o RNAt deslizando no ribossomo encontra os resíduos de adenina e libera o peptídeo formado.

Os aminoácidos são dados na Tabela 1.1.

1.3.3 O Código genético

No código genético universal todos os aminoácidos possíveis são representados por 61 códons, além dos 3 códons de terminação de cadeia. O código é altamente degenerado, como pode ser observado na Tabela 1.2, com vários códons correspondendo a um mesmo aminoácido.

1.4 Mutação

Para que os descendentes sobrevivam, o material genético parental deve ser transmitido de forma exata e inalterada à linhagem germinativa. Assim

Mutaç o

Amino�cido	Sigla	Amino�cido	Sigla
Alanina	Ala	Leucina	Leu
Arginina	Arg	Lisina	Lis
Asparagina	Asn	Metionina	Met
�cido Asp�rtico	Asp	Fenilalanina	Fen
Ciste�na	Cis	Prolina	Pro
Glutamina	Gln	Serina	Ser
�cido Glut�mico	Glu	Treonina	Tre
Glicina	Gli	Triptofano	Trp
Histidina	His	Tirosina	Tir
Isoleucina	Ile	Valina	Val

Tabela 1.1: Siglas dos amin cidos

	U	C	A	G	
U	Fen	Ser	Tir	Cis	U
	Fen	Ser	Tir	Cis	C
	Leu	Ser	PARE	PARE	A
	Leu	Ser	PARE	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Tre	Asn	Ser	U
	Ile	Tre	Asn	Ser	C
	Ile	Tre	Lis	Arg	A
	Met	Tre	Lis	Arg	G
G	Val	Ala	Asp	Gli	U
	Val	Ala	Asp	Gli	C
	Val	Ala	Glu	Gli	A
	Val	Ala	Glu	Gli	G

Tabela 1.2: Todos os amino cidos formados a partir das bases nitrogenadas

como a linhagem som tica de um organismo adulto que n o pode sofrer elevadas taxas de muta es de maneira a mudar as fun es proteicas das c lulas. Duas fontes importantes de muta o s o as falhas na replica o do DNA e as les es qu micas no material gen tico. A maquinaria enzim tica de replica o do DNA tenta compensar a incorpora o de nucleot deos incorretos por meio de um mecanismo de revis o de leitura, mas alguns erros escapam da detec o [26].

Existem três tipos de mutações que alteram o código genético:

- i) Mutaç o de sentido trocado: modifica um c don espec fico de um amino- cido para um c don de outro amino cido. Isso pode ocorrer pela mudan a de uma base nitrogenada por outra e conseq entemente a substitui o de um amino cido por outro na prote na.
- ii) Muta o sem sentido: altera o de uma base que provoca a forma o de um c don de termina o, formando um polipept deo incompleto.
- iii) Muta o de altera o de fase: inser o ou dele o de um ou alguns pares de bases que alteram completamente a leitura.

1.4.1 Mudan a na frequ ncia dos alelos

Para muitos eucariotos, as taxas de substitui es por base s o em torno de 10^{-9} por ano, o que mostra que a muta o sozinha   uma for a pequena na mudan a na frequ ncia dos alelos. Essas taxas s o maiores (entre 10^{-3} e 10^{-2} por base) em outros genomas, como em DNAs-mitocondriais de mam feros e RNA de v rus, os quais n o tem um tipo de mecanismo de reparo na replica o [25].

Supondo um modelo em que os alelos A podem trocar somente com os alelos a e n o vice versa, a mudan a na frequ ncia do alelo atrav s da muta o pode ser calculada por:

$$p_t = p_0(1 - \mu)^n \quad (1.4.1)$$

onde p_t   a frequ ncia do alelo A depois de t gera es e p_0 a frequ ncia inicial de A na popula o e μ a taxa de muta o. Uma vez que a taxa de muta o   t o pequena o termo $(1 - \mu)^n$ pode ser substituído por $e^{-\mu n}$ e a equa o pode ser reescrita como:

$$p_t = p_0 e^{-\mu n} \quad (1.4.2)$$

A partir dessa equa o pode-se observar que se a frequ ncia inicial do alelo A   1 e a taxa de muta o de A para a   10^{-5} por gera o, ent o depois de 100 gera es de muta es, a frequ ncia de A na popula o (p_t) diminuir  apenas para 0,999. Ser o necess rias quase 70000 gera es de muta es para reduzir a frequ ncia de A na popula o para 0,5.

A  nica maneira que a muta o pode afetar rapidamente a frequ ncia do alelo na popula o   quando ela ocorre persistentemente em um determinado locus. Uma vez que a probabilidade disso acontecer   muito pequena, essas muta es recorrentes envolvem mudan as gen ticas cruciais, incluindo grandes segmentos de cromossomos. Um exemplo de muta o recorrente   a talassemia, uma doen a gen tica comum em humanos.

Recombinação

Apesar da vagarosa influência da mutação na frequência dos genes na maioria dos casos, existe uma grande vantagem na ocorrência de uma pequena e finita taxa de mutação. Ela fornece uma fonte constante de novas variantes, necessárias para permitir que os seres vivos se ajustem aos ambientes físicos e biológicos em constante alteração. Se o material genético fosse perpetuado com fidelidade perfeita, a variação genética necessária para permitir a evolução seria perdida, e novas espécies, incluindo a espécie humana, não teriam surgido.

1.5 Recombinação

Quando os cromossomos homólogos são pareados antes da primeira divisão celular (meiose) ocorre uma permuta genética entre eles. Essa permuta física entre sequência de DNA nos cromossomos é conhecida como **crossing-over**. A frequência de crossing-over entre dois genes no mesmo cromossomo depende da distância física entre esses genes, quanto maior o afastamento entre eles maior a frequência de permuta.

A recombinação homóloga é um processo celular essencial, catalizado por enzimas sintetizadas e reguladas para esse fim. Além de gerar variações genéticas, a recombinação permite que as células recuperem sequências perdidas por lesões do DNA (molécula muito instável), através de substituição da região danificada por uma fita de DNA que não foi modificada de um cromossomo homólogo. A recombinação é conservativa, assim como a replicação, pois envolve a quebra e a religação do DNA. As etapas da migração de ramificações é mostrada na Figura 1.4.

A primeira etapa do processo necessita que uma das duas moléculas de DNA homólogas apresente uma quebra na fita dupla. As extremidades de DNA quebradas são processadas por enzimas que degradam o DNA, gerando segmentos de DNA de fita simples. Essas regiões de fita simples participam do pareamento com o parceiro de DNA homólogo. Uma vez ocorrido o pareamento as duas moléculas de DNA são unidas por uma estrutura ramificada no DNA, chamada de junção de Holliday [26].

1.6 Seleção natural

As mutações, como já foi citado, são responsáveis pela variação genética necessária para permitir a evolução. Um outro fator essencial para que ela ocorra é a **seleção natural**. A luta por sobrevivência faz com que os organismos compitam entre si, aqueles com genes que melhor se adaptam ao seu ambiente tem uma maior probabilidade de sobrevivência [15]. Além

disso, o que seleciona indivíduos geneticamente diferentes são suas características em relação a mortalidade, a fertilidade, a fecundidade, ao sucesso no acasalamento e a viabilidade de descendentes [25].

A maneira mais simples de pensar se um organismo está mais adaptado que outro é através da **aptidão**. Em termos de genética populacional aptidão é definida como a capacidade de um genótipo sobreviver e reproduzir. Isso é expresso em termos relativos, por exemplo, o heterozigoto Aa tem maior aptidão que os homozigotos AA e aa . A aptidão está relacionada com o ambiente, uma vez que um genótipo pode ser beneficiado em uma localização e deletério em outra.

Uma mutação pode ou não alterar o fenótipo de um indivíduo. Caso altere pode alterar também aptidão do mesmo. Na maioria dos casos ocorre **mutação deletéria** e ela será removida da população rapidamente por seleção negativa. Em casos mais raros pode ocorrer uma **mutação vantajosa**, aonde há um aumento da aptidão e uma seleção positiva, aonde o alelo mutado será favoravelmente fixado na população [15]. Um exemplo de seleção positiva envolve a evolução de resistência a antibióticos, drogas e inseticidas que ocorre com mosquitos submetidos a um controle através de DDT.

Em termos de evolução molecular a aptidão é expressa por um coeficiente de seleção, o qual mede o aumento do aptidão comparada com genótipos menos aptos na população. Por exemplo, se o coeficiente de seleção for denotado por s , e o seu valor for de 0,01 para um determinado genótipo, isso significa que esse genótipo tem 0,01% mais chance de sobrevivência que

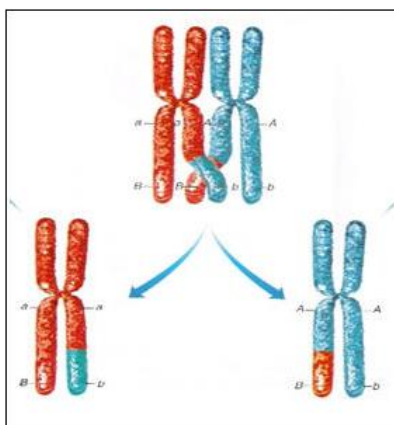


Figura 1.4: Crossing-over

Seleção natural

o genótipo menos favorecido. Quando se pensa em seleção natural deve-se considerar quatro casos possíveis:

- **Dominância:** o alelo A é dominante sobre a , o que significa que Aa tem a mesma aptidão que AA . Se os indivíduos que expressam A são mais aptos que os que expressam a , a seleção irá favorecer os genótipos AA e Aa . Veja a Tabela 1.3.

Genótipo	AA	Aa	aa
Aptidão	$1 + s$	$1 + s$	1

Tabela 1.3: Coeficiente de aptidão para um caso de dominância

- **Codominância:** neste caso, quando em heterozigotos, ambos os genes A e a são expressos, mas Aa tem aptidão intermediária à dos homozigotos. Veja a Tabela 1.4.

Genótipo	AA	Aa	aa
Aptidão	$1 + 2s$	$1 + s$	1

Tabela 1.4: Coeficiente de aptidão para um caso de codominância

- **Sobredominância:** o heterozigoto é favorecido sobre os homozigotos. O coeficiente de aptidão para o genótipo aa é $r < s$.

Genótipo	AA	Aa	aa
Aptidão	1	$1 + s$	$1 + r$

Tabela 1.5: Coeficientes de aptidão para um caso de sobredominância

- **Subdominância:** os homozigotos são favorecidos sobre os heterozigotos.

Genótipo	AA	Aa	aa
Aptidão	$1 + s$	1	$1 + r$

Tabela 1.6: Coeficientes de aptidão para um caso de subdominância

A partir desses dados, é possível mostrar como a seleção natural muda a frequência dos genes através de modelagens matemáticas.

Capítulo 2

Modelos Determinísticos

... nos assuntos em que o acaso governa tanto faz viver numa cidade de dez milhões de habitantes como numa aldeia de poucas centenas de moradores, só acontece o que tiver de acontecer.

José Saramago, *A Caverna*

2.1 Primeiras considerações

Como o conteúdo genético de uma população se transforma ao longo das gerações ?

Esta é a principal questão no âmbito da genética populacional.

Vamos considerar as seguintes

Hipóteses 2.1 (Modelo 1).

H1 - A população possui N indivíduos diplóides;

H2 - A análise é feita sobre um locus gênico;

H3 - Dois alelos A e a são observados para esse locus.

Os possíveis genótipos para os indivíduos desta população são: AA , Aa , aa

Sendo $\#AA$, $\#Aa$, $\#aa$ o número de genótipos AA , Aa e aa presentes, respectivamente, na população, e $\#A$ e $\#a$ o número de genes A e a na

Modelos Determinísticos

população, respectivamente, então

$$\begin{aligned}\#A &= 2(\#AA) + \#Aa \\ \#a &= 2(\#aa) + \#Aa\end{aligned}\quad (2.1.1)$$

Assim, o número de indivíduos é $N = \#AA + \#Aa + \#aa$ e o número de genes é $2N = 2\#AA + 2\#Aa + 2\#aa = \#A + \#a$.

Sejam

$$\begin{aligned}f &= \text{freqüência de } A = \frac{\#A}{\#A + \#B} \\ g &= \text{freqüência de } B = 1 - f\end{aligned}\quad (2.1.2)$$

e também

$$\begin{aligned}x &= \text{freqüência de } AA = \frac{\#AA}{\#AA + \#Aa + \#aa} \\ y &= \text{freqüência de } Aa = \frac{\#Aa}{\#AA + \#Aa + \#aa} \\ z &= \text{freqüência de } aa = 1 - x - y.\end{aligned}\quad (2.1.3)$$

Podemos expressar p, q de forma *única* como funções de x, y, z : de fato

$$\begin{aligned}f &= \frac{2(\#AA) + \#Aa}{2(\#AA) + 2\#Aa + 2\#aa} = x + \frac{1}{2}y \\ g &= \frac{2(\#aa) + \#Aa}{2(\#AA) + 2\#Aa + 2\#aa} = z + \frac{1}{2}y\end{aligned}\quad (2.1.4)$$

No entanto, note que se quisermos expressar x, y, z como funções de f, g teremos múltiplas formas. Por exemplo, se tivermos 6 genes A e 4 genes a numa população de 5 indivíduos, então $f = \frac{3}{5}$ e $g = \frac{2}{5}$ e teremos as seguintes possibilidades para os genótipos dadas na Tabela 2.1 :

População	(x, y, z)
AA, Aa, Aa, Aa, Aa	$(1/5, 4/5, 0)$
AA, AA, aa, Aa, Aa	$(2/5, 2/5, 1/5)$
AA, AA, AA, aa, aa	$(3/5, 0, 2/5)$

Tabela 2.1: Exemplo de Possibilidades de freqüências de Genótipos

Devido a essa multiplicidade de combinações, é comum fazer-se a seguinte hipótese adicional:

H4 - *O número esperado de emparelhamentos de um genótipo 1 com um genótipo 2 é proporcional ao produto da freqüência do genótipo 1 com a freqüência do genótipo 2.*

Então temos:

$$\begin{aligned}x &= f^2 \\ y &= 2fg \\ z &= g^2\end{aligned}\quad (2.1.5)$$

Modelo de Hardy-Weinberg

o que significa que “extraímos um indivíduo AA com frequência f^2 , extraímos um indivíduo Aa com frequência $2fg$, e extraímos um indivíduo aa com frequência g^2 ”. Ou ainda, numa linguagem probabilística, para formarmos um indivíduo AA devemos selecionar o gene A duas vezes, como A tem frequência f , então a probabilidade de formar AA é f^2 . E da mesma forma para os outros genótipos.

Observação 2.1. *Note que para a hipótese **H4** ser aplicável, a população deve possuir potencialmente infinitos indivíduos. Assim sob a hipótese **H4**, a hipótese **H1** não pode ser válida.*

Exercício 2.1. *Quais outras hipóteses estão subentendidas para se obter as equações (2.1.5) ?*

2.2 Modelo de Hardy-Weinberg

As equações apresentadas na seção anterior indicam apenas o estado da população num dado momento, sob aquele conjunto de hipóteses, exibindo um *retrato* das frequências gênicas e genótípicas da população num instantâneo. Elas não respondem à pergunta formulada no início do capítulo. Objetivando responder àquela questão, considere o conjunto de hipóteses a seguir.

Hipóteses 2.2 (Modelo de Hardy-Weinberg).

HW1 - *O número de indivíduos na população é infinito.*

HW2 - *A análise é feita sobre um locus gênico;*

HW3 - *Dois alelos A e a são observados para esse locus.*

HW4 - *A fertilidade e a sobrevivência são independentes do genótipo.*

HW5 - *Não há mutação.*

Seja f_n a frequência do gene A na geração n e x_n, y_n, z_n as frequências dos genótipos AA, Aa e aa na geração n respectivamente. Vamos considerar um esquema de reprodução segundo a hipótese

HW6 (*Random mating*): a frequência dos genótipos dos indivíduos na geração n são obtidos proporcionalmente ao produto das frequências dos genótipos na geração $n - 1$.

Ou seja, a hipótese **HW6** é uma adaptação da hipótese **H4** para imbutir *dinâmica temporal* no modelo.

Então

$$\begin{aligned} f_n &= x_{n-1} + \frac{1}{2}y_{n-1} \\ &= f_{n-1}^2 + \frac{1}{2}f_{n-1}g_{n-1} \\ &= p_{n-1} \end{aligned} \tag{2.2.6}$$

Por um raciocínio análogo, $g_n = g_{n-1}$. Sendo f_0, g_0 as frequências na geração inicial, então $f_n = f_0$ e $g_n = g_0$, para todo $n \in \mathbb{N}$.

Além disso,

$$x_n = f_n^2 = f_0^2 \quad (2.2.7)$$

$$y_n = 2f_n g_n = 2f_0 g_0 \quad (2.2.8)$$

$$z_n = g_n^2 = g_0^2 \quad (2.2.9)$$

Este resultado básico é conhecido como **modelo de Hardy-Weinberg**: “As frequências dos alelos não se alteram de uma geração a outra, e a partir da primeira geração a frequência dos genótipos também não se altera.”

Note que este resultado é obtido por meio de hipóteses muito restritivas: a população deve ser infinita, não pode haver pressão seletiva, nem mutação, nem migração, etc. Se alguma dessas hipóteses não for válida, os resultados do modelo não vão se aplicar.

A frequência total de homozigotos, sejam eles AA ou aa é

$$G = f^2 + (1 - f)^2. \quad (2.2.10)$$

G é denominado **homozigosidade**. A **heterozigosidade** é definida por

$$H = 1 - G = 2f(1 - f). \quad (2.2.11)$$

Para populações que satisfazem a hipótese de *random mating*, a heterozigosidade é igual à frequência de heterozigotos. Veja ainda que a definição de heterozigosidade utiliza apenas as frequências dos alelos, e não a dos genótipos. Assim a heterozigosidade serve para medir níveis de variações de populações que não estão de acordo com a hipótese de *random mating* de teorema de Hardy-Weinberg.

Exercício 2.2. *Como se transformam G e H sob as hipóteses do modelo de Hardy-Weinberg?*

2.3 Seleção Natural

Vamos, agora, incluir a “luta pela sobrevivência” no modelo. Para isso, devemos descartar a hipótese **HW4**.

Vamos continuar supondo **HW1, HW2, HW3, HW5, HW6**.

Precisaremos utilizar a noção de aptidão de um indivíduo.

Aptidão absoluta: “número de cópias de cada gene que um indivíduo de um certo genótipo espera contribuir para o conjunto de genes da geração

Seleção Natural

seguinte.”

Aptidão relativa: “razão entre a aptidão absoluta para a aptidão absoluta de um genótipo de referência”.

Sejam w_x, w_y e w_z as aptidões relativas dos indivíduos de genótipos AA , Aa e aa respectivamente, e suponhamos que elas permaneçam constantes ao longo de todas as gerações.

As razões dos genótipos na geração n são

$$w_x f_n^2 : 2w_y f_n g_n : w_z g_n^2 \quad (2.3.12)$$

A geração $n + 1$ será composta por

$$f_{n+1} = \frac{(w_x f_n + w_y g_n) f_n}{w_x f_n^2 + 2w_y f_n g_n + w_z g_n^2} \quad (2.3.13)$$

ou ainda

$$f_{n+1} = f_n + h(f_n) \quad (2.3.14)$$

onde

$$h(f_n) = f_n g_n \frac{(w_x - w_y) f_n + (w_y - w_z) g_n}{w_x f_n^2 + 2w_y f_n g_n + w_z g_n^2} \quad (2.3.15)$$

Esta é a **Equação de Fisher-Haldane-Wright**.

Exercício 2.3. *Encontre os pontos estacionários da equação (2.3.14), ou seja, os valores de f_n para os quais $h(f_n) = 0$.*

Para referência futura, façamos as seguintes definições:

$$\begin{aligned} w_f &= \frac{w_x f_0^2 + w_y f_0 g_0}{f_0^2 + f_0 g_0} = w_x f_0 + w_y g_0 \\ w_g &= \frac{w_y f_0 g_0 + w_z g_0^2}{f_0 g_0 + g_0^2} = w_y f_0 + w_z g_0 \\ \bar{w} &= \frac{w_x f_0^2 + 2w_y f_0 g_0 + w_z g_0^2}{f_0^2 + 2f_0 g_0 + g_0^2} = f_0 w_f + g_0 w_g \end{aligned} \quad (2.3.16)$$

Na dependência das relações de dominância entre os genes A e a , o comportamento de f_n com n irá variar. Vejamos, a seguir, alguns casos.

2.3.1 Dominância

Considere A dominante. Então AA e Aa tem a mesma aptidão e estamos supondo que a aptidão de AA e Aa é maior que a de aa . Veja a Tabela 2.2. Substituindo os valores de aptidão na equação (2.3.14) obtemos

Genótipo	AA	Aa	aa
Aptidão	1 + s	1 + s	1

Tabela 2.2: Dominância

$$f_{n+1} = f_n + sf_n g_n \frac{g_n}{1 + s(f_n^2 + 2f_n g_n)} \quad (2.3.17)$$

Para $f_0 \approx 0$ e $g_0 \approx 1$, $f_{n+1} \approx f_n + sf_n$ e temos

$$f_n \approx (1 + s)^n f_0$$

isto é, para uma frequência baixa de A , f_n tem um aumento geométrico. Este resultado é esperado, pois quando há poucos genes A e muitos a , o gene A tende a estar presente nos heterozigotos Aa e praticamente inexistem indivíduos AA . Como a aptidão de Aa é maior do que a de aa , e como Aa não compete com AA , então a frequência de A aumenta rapidamente.

Para $f_0 \approx 1$ e $g_0 \approx 0$, seja $u = \frac{f}{g}$, temos

$$\frac{f_{n+1}}{g_{n+1}} - \frac{f_n}{g_n} \approx sf_n \frac{g_n}{1 + s(f_n^2 + 2f_n g_n)}$$

ou

$$u_{n+1} - u_n \approx \frac{s}{s + 1}$$

ou

$$u_n = u_0 + n \frac{s}{s + 1}.$$

Isto significa que para uma frequência de A já próxima de 1, o crescimento de f_n se torna linear, bem mais lento que no caso $f_0 \approx 0$. Em outras palavras, quando A aparece numa frequência muito alta, a maioria dos indivíduos é do tipo AA , os quais competem entre si para deixar descendentes e ainda eliminar os restantes dos genes a . Por isso o crescimento da frequência do gene A à medida que se aproxima de 1 se torna lenta.

2.3.2 Recessividade

Considere A recessivo. Supomos que a aptidão de aa e Aa sejam iguais e menores que a de AA . Veja a Tabela 2.3. Da equação (2.3.14) obtemos

$$f_{n+1} = f_n + sf_n g_n \frac{f_n}{1 - sf_n^2} \quad (2.3.18)$$

Exercício 2.4. Analise o comportamento de f_n próximo de 0 e de 1.

Seleção Natural

Genótipo	AA	Aa	aa
Aptidão	1 + s	1	1

Tabela 2.3: Recessividade

2.3.3 Caso aditivo - codominância

Neste caso, a aptidão do heterozigoto é intermediária à dos homozigotos. Veja a Tabela 2.4. Da equação (2.3.14) obtemos

Genótipo	AA	Aa	aa
Aptidão	1 + 2s	1 + s	1

Tabela 2.4: Codominância

$$f_{n+1} = f_n + \frac{sf_n g_n}{1 + sf_n} \quad (2.3.19)$$

Exercício 2.5. *Faça gráficos comparativos de f_n e apreenda seu comportamento próximo de 0 e de 1.*

2.3.4 Seleção fraca

No caso em que o fator de seleção é fraco ($s \ll 1$), podemos aproximar continuamente a equação discreta de f_n (2.3.14). Para isso, considere:

$$\begin{aligned} w_x &= 1 + O(s) \\ w_y &= 1 + O(s) \\ w_z &= 1 + O(s) \end{aligned} \quad (2.3.20)$$

ou mais especificamente, para $s \ll 1$, e h, k constantes

$$w_x = 1 + hs \quad (2.3.21)$$

$$w_y = 1 + ks \quad (2.3.22)$$

$$w_z = 1 \quad (2.3.23)$$

então, substituindo em (2.3.14) temos

$$f_{n+1} - f_n = f_n g_n [(h - k)s f_n + k s g_n] \quad (2.3.24)$$

onde aproximamos o denominador $w_x f_n^2 + 2w_y f_n g_n + w_z g_n^2$ por 1. Agora fazendo a aproximação $f_{n+1} - f_n \approx \dot{f}$, obtemos

$$\dot{f} = sfg[(h - k)f + kg]. \quad (2.3.25)$$

A solução da equação diferencial (2.3.25) é obtido por uma integração simples

$$t = \frac{1}{s} \int_{f_0}^{f_1} \frac{df}{f(1-f)[(h-k)f + k(1-f)]} \quad (2.3.26)$$

Exercício 2.6. Calcule (2.3.26). *Sugestão: utilize frações parciais.*

Para o caso aditivo, sendo A vantajoso, $k = 1$, $h = 2$, temos

$$\frac{df}{dt} = sf(1-f). \quad (2.3.27)$$

Observação 2.2. Uma versão espacial para a equação (2.3.27) é

$$\frac{\partial f}{\partial t} = sf(1-f) + D \frac{\partial^2 f}{\partial \xi^2}. \quad (2.3.28)$$

Aqui, $f(t, \xi)$ representa a frequência do gene tanto no tempo quanto no espaço, e D é um coeficiente de difusão representando o movimento aleatório dos genes no espaço.

2.4 Mutações

A evolução atua nas mutações que ocorrem na linha germinativa de um indivíduo, por erros de transcrição, agentes mutagênicos, etc. Em geral as mutações são mais deletérias do que benéficas. Elas ajudam a manter um suplemento de variação genética sobre a qual atua a seleção.

Para construir um modelo, devemos abandonar a hipótese **HW5**, mas vamos continuar supondo **HW1, HW2, HW3, HW4, HW6**.

Suponha que o gene A se transforme em a com probabilidade u , e que a se transforme em A com probabilidade v . Se f_n é a frequência de A na geração n , então na geração $n + 1$, $(1 - u)f_n + vg_n$ serão A :

$$f_{n+1} = (1 - u)f_n + vg_n \quad (2.4.29)$$

Fazendo $f^* = \frac{v}{u+v}$ temos

$$f_{n+1} - f^* = (1 - u - v)(f_n - f^*) \quad (2.4.30)$$

Resolvendo fornece

$$f_n = f^* + (f_0 - f^*)(1 - u - v)^n. \quad (2.4.31)$$

Mutação

Para $u + v < 1$, temos

$$\lim_{n \rightarrow \infty} = f^*. \quad (2.4.32)$$

A população, na presença de mutação, tende a um estado de equilíbrio onde as frequências dos genes são não-nulas.

Exercício 2.7. *Mostre que f^* é ponto estacionário da equação (2.4.29).*

2.4.1 Equilíbrio entre seleção e mutação

O fator de seleção favorece determinado genótipo em detrimento de outro e leva toda a população, com o passar do tempo a se tornar homogênea. Já a mutação atua no sentido de manter a variedade genética da população. *Qual o equilíbrio entre essas duas forças ?*

Vamos considerar um modelo onde nem **HW4** e nem **HW5** são satisfeitos, ou seja, um modelo com seleção natural e mutação. Ainda supomos **HW1, HW2, HW3, HW6**.

Defina f, f' como sendo as frequências do gene A na geração corrente e na subsequente. Então de (2.3.12) e de (2.4.29)

$$f' = (1 - u) \frac{w_f}{\bar{w}} f + v \frac{w_g}{\bar{w}} g \quad (2.4.33)$$

onde usamos (2.3.16).

Seja $\alpha_f = w_f - \bar{w}$ que é a diferença de aptidão média do gene A em relação à média populacional. Seja ainda $\delta f = f' - f$, a variação da frequência entre as gerações. Então

$$\delta f = \frac{\alpha_f f}{\bar{w}} - u \frac{w_f}{\bar{w}} f + v \frac{w_g}{\bar{w}} g. \quad (2.4.34)$$

Como

$$\delta f_{\text{sel}} = \frac{\alpha_f f}{\bar{w}}$$

é a contribuição da seleção natural para a variação da frequência e como

$$\delta f_{\text{mut}} = -u \frac{w_f}{\bar{w}} f + v \frac{w_g}{\bar{w}} g$$

é a contribuição da mutação, temos

$$\delta f = \delta f_{\text{sel}} + \delta f_{\text{mut}} \quad (2.4.35)$$

Adequando os parâmetros de seleção e mutação, os efeitos de ambos fatores poderão se cancelar e teremos um estado de estacionário.

Modelos Determinísticos

Exercício 2.8 (Britton). *Suponha que os parâmetros de mutação são muito pequenos se comparados com os de seleção, os quais também são pequenos. Desconsidere termos de segunda ordem de quantidades pequenas. Suponha que a seleção natural é negativa para um gene deletério. Se a for recessivo, mostre que seu valor estacionário é*

$$f^* = \sqrt{\frac{v}{s}}.$$

Notas

As referências para este capítulo são BRITTON [1], capítulo 4; GILLESPIE [14], capítulos 1 e 3, GRAUR & LI [15], capítulo 2.

Capítulo 3

Modelos Probabilísticos

On the one hand, we expect from chance that it creates random fluctuations, and on the other hand that it averages them out. This relation between mean and dispersion is not at all easy to grasp.

Karl Sigmund, *Games of Life*

3.1 Modelo de Wright-Fisher

Um fator importante que produz flutuações aleatórias nas frequências dos genes é a amostragem aleatória dos gametas durante o processo de reprodução, numa população de tamanho finito. Em vista da quantidade potencial de geração de gametas ser muito grande em comparação ao número de indivíduos que são efetivamente gerados, ocorre uma *amostragem* a partir do conjunto de gametas para gerar os indivíduos. A estocasticidade produzida por esse efeito aleatório é denominada **deriva genética aleatória**. Esse fenômeno pode ser modelado por aquele que é um dos modelos mais simples e mais largamente utilizados em genética de populações, o **modelo de Wright-Fisher**. A hipótese de que a distribuição dos genes na geração $t + 1$ somente depende da distribuição na geração t é crucial. Essa hipótese foi utilizada implicitamente por Fisher e explicitamente por Wright na década de 30 (ver EWENS [9]).

Hipóteses 3.1 (Modelo de Wright-Fisher).

WF1 - *O número de indivíduos na população é fixo e igual a N .*

Modelos Probabilísticos

WF2 - A análise é feita sobre um locus gênico;

WF3 - Dois alelos A e a são observados para esse locus.

WF4 - A fertilidade e a sobrevivência são independentes do genótipo.

WF5 - Não há mutação.

WF6 - Os indivíduos são diplóides

WF7 - As gerações não se sobrepõem.

Seja f_n a frequência do gene A na geração n . Alternativamente iremos trabalhar também com a variável X_n que representa o número de genes A presentes na população na geração n . A relação entre f_n e X_n é

$$f_n = \frac{X_n}{2N}. \quad (3.1.1)$$

Os indivíduos presentes na geração n podem gerar uma infinidade de gametas dos tipos A e a . A proporção de A e a no conjunto dos gametas é igual à proporção de A e a na população. Dessa forma, retirando uma amostra de $2N$ genes do conjunto de gametas estabelecemos qual será a composição da população no instante seguinte.

Para calcular a probabilidade de que a amostra de tamanho $2N$ contenha exatamente j alelos do tipo A , devemos considerar:

(i) o número total de amostras que contém exatamente j alelos do tipo A corresponde à uma permutação dos $2N$ genes com A repetido j vezes e a repetido $(2N - j)$ vezes, ou seja,

$$\frac{(2N)!}{j!(2N - j)!};$$

(ii) cada gene A tem probabilidade f_n de estar na amostra, pois esta é a proporção dele na população; cada gene a é sorteado com probabilidade $(1 - f_n)$, portanto retiramos j genes A e $(2N - j)$ genes a com probabilidade $f_n^j (1 - f_n)^{2N - j}$.

Logo, a probabilidade da amostra conter exatamente j genes A para a geração n , dado que a frequência do gene A na geração n é $f_n = \frac{i}{2N}$ é calculada por

$$Prob \left[f_{n+1} = \frac{j}{2N} \mid f_n = \frac{i}{2N} \right] = \frac{(2N)!}{j!(2N - j)!} \left(\frac{i}{2N} \right)^j \left(1 - \frac{i}{2N} \right)^{2N - j}. \quad (3.1.2)$$

Alternativamente, podemos escrever

$$Prob [X_{n+1} = j \mid X_n = i] = \frac{(2N)!}{j!(2N - j)!} \left(\frac{i}{2N} \right)^j \left(1 - \frac{i}{2N} \right)^{2N - j}. \quad (3.1.3)$$

Modelo de Wright-Fisher

O valor de X_{n+1} dado X_n , isto é, $X_{n+1}|X_n$ segue uma distribuição binomial.

Notemos que a hipótese de Fisher e Wright sobre a obtenção da geração $n+1$ deixa implícito que toda *informação* necessária para gerar os indivíduos da geração $n+1$ está contida na geração n .

O valor esperado de f_{n+1} dado f_n é

$$E[f_{n+1}|f_n] = f_n. \quad (3.1.4)$$

A variância de f_{n+1} dado f_n é

$$Var[f_{n+1}|f_n] = \frac{f_n(1-f_n)}{2N} \quad (3.1.5)$$

A equação (3.1.4) mostra que as frequências dos alelos na população se mantêm constante, em média. Porém, por causa das flutuações aleatórias, em qualquer população dada, a frequência não se manterá constante.

Exercício 3.1. Utilize propriedades do somatório para mostrar as expressões (3.1.4) e (3.1.5). Lembre-se que, por definição, que se X é uma variável aleatória discreta assumindo os valores x_1, x_2, \dots, x_r com probabilidades p_1, p_2, \dots, p_r , respectivamente, então

$$E[X] = \sum_{k=1}^r x_k p_k. \quad (3.1.6)$$

A variância de X é definida por

$$Var[X] = E\{[X - E(X)]^2\}. \quad (3.1.7)$$

3.1.1 Probabilidade de fixação de A

O modelo de Wright-Fisher sem mutação e sem seleção prevê que eventualmente um dos alelos se fixará na população, isto é, a deriva genética aleatória torna a população cada vez mais homogênea [5, 29].

De fato, seja X_n o número de genes A presentes na geração n , que satisfaz (3.1.1). Seja o tempo aleatório $\tau = \min\{n : X_n = 0 \text{ ou } X_n = 2N\}$, que é o tempo que A leva para se fixar na população ou desaparecer totalmente.

Defina

$$E_i[X_t] = E[X_t|X_0 = i]$$

Então

$$E_i[X_\tau] = E_i[X_0] = i.$$

Como $X_\tau = 0$ ou $2N$ temos

$$i = E_i[X_\tau] = 0P_i[X_\tau = 0] + 2NP_i[X_\tau = 2N]. \quad (3.1.8)$$

Donde

$$P_i[X_\tau = 2N] = \frac{i}{2N}.$$

Isto é, a probabilidade do gene A se fixar na população é igual a sua frequência inicial.¹

Para medir quanto tempo levará até que ocorra a fixação de um gene, podemos determinar o **estado de heterozigose** \mathcal{H}_n da população [5], que é a probabilidade de que dois genes tomados da população *sem reposição* no instante n sejam diferentes, ou seja,

$$\mathcal{H}_n = \frac{2X_n(2N - X_n)}{2N(2N - 1)}. \quad (3.1.9)$$

Exercício 3.2. *Mostre que o estado de heterozigose \mathcal{H}_n é quase igual à heterozigosidade H , em vista de*

$$H = \left(1 - \frac{1}{2N}\right) \mathcal{H}. \quad (3.1.10)$$

O valor esperado de \mathcal{H}_n é

$$E[\mathcal{H}_n] = \left(1 - \frac{1}{2N}\right)^n E[\mathcal{H}_0] \quad (3.1.11)$$

o que mostra que o estado de heterozigose decresce geometricamente a zero, em média.

Note, ainda, que quando x é pequeno, $(1 - x) \approx e^{-x}$, então se N for grande

$$E[\mathcal{H}_n] \approx e^{-\frac{n}{2N}} \mathcal{H}_0 \quad (3.1.12)$$

mostrando que o estado de heterozigose decai para 0 a uma taxa exponencial, à medida que $\frac{n}{2N} \rightarrow \infty$.

Exercício 3.3. *Mostre que a heterozigosidade H satisfaz*

$$E[H_n] = \left(1 - \frac{1}{2N}\right)^n E[H_0] \quad (3.1.13)$$

e então pode ser aproximada por

$$E[H_n] = E[H_0] e^{-\frac{n}{2N}}. \quad (3.1.14)$$

¹Um forma intuitiva de obter a probabilidade de fixação de um gene, é notar que eventualmente cada gene na população é descendente de um único gene na geração inicial. A probabilidade de que esse gene seja A é simplesmente sua frequência inicial.

Modelo de Wright-Fisher

Exercício 3.4. Calcule o número de gerações necessárias para reduzir \mathcal{H}_0 pela metade.

Seja \mathcal{G} a probabilidade de que dois alelos são idênticos por estado, apesar de diferentes por origem (i.e. os genes são tomados sem reposição), denominado **estado de homozigose**. Temos, então, que

$$\mathcal{G} = 1 - \mathcal{H} \quad (3.1.15)$$

Exercício 3.5. Mostre que a homozigosidade G é quase igual ao estado de homozigose \mathcal{G} , valendo a expressão:

$$G = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) \mathcal{G}. \quad (3.1.16)$$

3.1.2 Mutaçãõ

A deriva genética elimina a variação genética de uma população. Para restaurar a variação genética, entra em cena o fator de **mutaçãõ**. Isto significa abdicar da hipótese **WF5**.

Para modelar a mutaçãõ no processo de Wright-Fisher, podemos supor que com uma probabilidade u o gene A se transforma em a , enquanto com probabilidade v o gene a se transforma em A . Partindo da n -ésima geraçãõ, cada gene escolhido antes de entrar na geraçãõ $n + 1$, pode sofrer mutaçãõ, assim a probabilidade de escolher um gene A para a populaçãõ no tempo $n + 1$ quando há i genes do tipo A na geraçãõ n é

$$p_i = \frac{i}{2N}(1 - u) + \frac{2N - i}{2N}v \quad (3.1.17)$$

e a probabilidade de haja j genes A na geraçãõ $n + 1$, dado que há i genes A na geraçãõ n fica

$$Prob_{\text{mut}}(X_{n+1} = j | X_n = i) = P_{ij} = \frac{(2N)!}{(i)!(2N - i)!} p_i^j (1 - p_i)^{2N - j}. \quad (3.1.18)$$

Exercício 3.6. Calcule o valor esperado e a variância de X_{n+1} , dado X_n , de acordo com distribuiçãõ de probabilidade (3.1.18), mostrando que

$$\begin{aligned} E[X_{n+1} | X_n = i] &= (1 - 2Nu) \frac{i}{2N} + 2Nv \left(1 - \frac{i}{2N}\right) \\ Var[X_{n+1} | X_n = i] &= \frac{i(2N - i)}{2N} \end{aligned} \quad (3.1.19)$$

Modelos Probabilísticos

O modelo não possui estado absorvente, ou seja, nenhum gene se fixa, pois $P_{ij} > 0$ para todo i, j .

De fato, considere a composição da população no limite quando $n \rightarrow \infty$, dada por X_∞ . Vamos calcular a $E[X_\infty]$ e $Var[X_\infty]$.

Temos que $E[X_{n+1}] = (1 - u)E[X_n] + (2N - E[X_n])v$. Como estamos procurando um estado limite de X_n , então para n grande devemos ter $E[X_n] = E[X_{n+1}] = x$, então

$$x = (1 - u)x + (2N - x)v$$

ou seja

$$x = \frac{2Nv}{v + u}.$$

Fazendo $\rho = \frac{v}{v+u}$ temos:

$$E[X_{n+1} - 2N\rho] = (1 - v - u)E[X_n - 2N\rho].$$

Então, se $0 < u + v < 2$ temos

$$E[X_n] \rightarrow 2N\rho \text{ quando } n \rightarrow \infty.$$

Logo

$$E[X_\infty] = \frac{2Nv}{v + u}. \quad (3.1.20)$$

Ou seja, em média, a composição limite da população consta de $\frac{v}{v+u}$ genes A e $\frac{u}{v+u}$ genes a .

Antes de calcular a variância, vamos entender o processo de coalescência dos genes.

Dois genes são **idênticos por descendência** se suas linhagens coalescem antes que uma mutação afete uma ou outra linhagem.

Seja $\mu = u + v$ a probabilidade de mutação em uma geração, então a probabilidade de que dois genes sejam iguais por descendência é

$$\rho \approx \frac{1}{1 + 4N\mu}. \quad (3.1.21)$$

De fato, uma mutação em cada linhagem pode ocorrer com probabilidade 2μ ou um evento de coalescência pode ocorrer com probabilidade $\frac{1}{2N}$. A probabilidade ρ de mutação antes de coalescência após um ciclo satisfaz a

$$\rho = 2\mu + (1 - 2\mu)\left(1 - \frac{1}{2N}\right)\rho \quad (3.1.22)$$

Modelo de Wright-Fisher

pois se nenhum evento ocorre, tudo começa outra vez. Assim

$$\rho = \frac{2\mu}{2\mu + \frac{1}{2N} - 2\mu\frac{1}{2N}}. \quad (3.1.23)$$

Ignorando a possibilidade de mutação e coalescência ocorrer num mesmo instante, ou seja, para $2\mu\frac{1}{2N} \approx 0$, temos

$$\rho \approx \frac{2\mu}{2\mu + \frac{1}{2N}}, \quad (3.1.24)$$

o que dá o resultado desejado.

Vamos então calcular a variância de X_∞ .

Seja $X_\infty = \sum_{i=1}^{2N} \mathbf{I}_i$, onde \mathbf{I}_i indica se o i -ésimo gene é A , caso em que $\mathbf{I}_i = 1$, ou não é A , caso em que $\mathbf{I}_i = 0$. Temos

$$X_\infty^2 = \sum_{i=1}^{2N} \sum_{j=1}^{2N} \mathbf{I}_i \mathbf{I}_j,$$

assim

$$E[X_\infty^2] = 2NP(\mathbf{I}_1 = 1) + 2N(2N - 1)P(\mathbf{I}_1 = 1, \mathbf{I}_2 = 1). \quad (3.1.25)$$

Como, de (3.1.20), $P(\mathbf{I}_1 = 1) = \frac{v}{v+u}$ e usando 3.1.22

$$P(\mathbf{I}_1 = 1, \mathbf{I}_2 = 1) = \frac{1}{1 + 4N(v+u)} \frac{v}{v+u} + \frac{4N(v+u)}{1 + 4N(v+u)} \left(\frac{v}{v+u} \right)^2, \quad (3.1.26)$$

daí

$$E[X_\infty^2] = \left\{ 2N + 2N(2N - 1) \left[\frac{1}{1 + 4N(v+u)} \frac{v}{v+u} + \frac{4N(v+u)}{1 + 4N(v+u)} \right] \right\} \times \left(\frac{v}{v+u} \right)^2. \quad (3.1.27)$$

Segue portanto que

$$Var(X_\infty^2) = \left(2N + \frac{2N(2N - 1)}{1 + 4N(v+u)} \right) \left(\frac{vu}{v+u} \right)^2. \quad (3.1.28)$$

3.1.3 Seleção

Considerando diferenças seletivas entre os possíveis genótipos, devemos descartar a hipótese **WF4**. Suponhamos que cada genótipo têm as seguintes aptidões relativas:

$$\begin{aligned} AA &\rightarrow \sigma_{AA} = 1 + s \\ Aa &\rightarrow \sigma_{Aa} = 1 + ss^* \\ aa &\rightarrow \sigma_{aa} = 1 \end{aligned}$$

sendo σ_{AA} , σ_{Aa} e σ_{aa} a viabilidade dos indivíduos AA , Aa e aa se reproduzirem, respectivamente, s e s^* valores pequenos, da ordem de 1% (exceto em casos letais), teremos novas probabilidades na escolha dos genes para a geração seguinte. Para o caso de serem os genes codominantes, temos $s^* = \frac{1}{2}$.

Se houver i genes A na geração atual, denotaremos η_i a probabilidade do gene A ser selecionado para a próxima geração. O genótipo AA contribui com 2 genes, o genótipo Aa contribui com 1 apenas (mas não nos esqueçamos de que $Aa = aA$) assim

$$\eta_i = \frac{\sigma_{AA}i^2 + \frac{1}{2}2\sigma_{Aa}i(2N - i)}{\sigma_{AA}i^2 + 2\sigma_{Aa}i(2N - i) + \sigma_{aa}(2N - i)^2}. \quad (3.1.29)$$

A transição do estado i para o estado j terá a seguinte probabilidade

$$Prob_{\text{sel}}(X_{n+1} = j | X_n = i) = P_{ij} = \binom{2N}{j} (\eta_i)^j (1 - \eta_i)^{2N-j}. \quad (3.1.30)$$

Exercício 3.7. Calcule o valor esperado e a variância de X_{n+1} , dado X_n , de acordo com distribuição de probabilidade (3.1.30), mostrando que

$$\begin{aligned} E[X_{n+1} | X_n = i] &= 2Ns \frac{i}{2N} (1 - \frac{i}{2N}) [\frac{i}{2N} + s^*(1 - 2\frac{i}{2N})] + \frac{i}{2N} \\ Var[X_{n+1} | X_n = i] &= \frac{i(2N-i)}{2N} \end{aligned} \quad (3.1.31)$$

3.1.4 Mutação e Seleção

Por fim, se permitirmos que haja seleção e mutação de A para a com probabilidade u , e de a para A com probabilidade v , devemos descartar as hipóteses **WF4** e **WF5**, obtendo a frequência *alterada* de escolha do gene A :

$$\eta_i^* = \underbrace{(1-u)\eta_i}_{\text{fração de } A \text{ que não sofre mutação}} + \underbrace{v(1-\eta_i)}_{\text{fração de } a \text{ que vira } A}$$

Modelo de Wright-Fisher

e a probabilidade de transição do estado i para o estado j

$$Prob_{\text{mut,sel}}(X_{n+1} = j | X_n = i) = P_{ij} = \binom{2n}{j} (\eta_i^*)^j (1 - \eta_i^*)^{2n-j}. \quad (3.1.32)$$

Exercício 3.8. Calcule o valor esperado e a variância de X_{n+1} , dado X_n , de acordo com distribuição de probabilidade (3.1.32), mostrando que

$$\begin{aligned} E[X_{n+1} | X_n = i] &= 2Ns \frac{i}{2N} \left(1 - \frac{i}{2N}\right) \left[\frac{i}{2N} + s^* \left(1 - 2\frac{i}{2N}\right)\right] \\ &\quad + (1 - 2Nu) \frac{i}{2N} + 2Nv \left(1 - \frac{i}{2N}\right) \\ Var[X_{n+1} | X_n = i] &= \frac{i(2N-i)}{2N} \end{aligned} \quad (3.1.33)$$

3.1.5 Aproximação por difusão

Quanto maior N , mais complexo se torna fazer cálculos com as expressões (3.1.18), (3.1.30) e (3.1.32) e obter respostas simples a questões relevantes em genética populacional [28]. Por isso nos empenhamos em apresentar um modelo de difusão no intervalo $[0, 1]$ que aproxime o processo $f_n = \frac{X_n}{2N}$ para N grande [9].

Seja $x \in [0, 1]$ a fração (frequência) de genes A na população com N indivíduos. Utilizaremos as seguintes definições:

$$\alpha_s = 2Ns, \quad \alpha_u = 2Nu, \quad \alpha_v = 2Nv$$

sendo s, u e v como na seção anterior. Devido ao modelo binomial, para $x = \frac{i}{2N}$ e $x + \delta x = \frac{j}{2N}$, temos

$$\begin{aligned} E[\delta x | x] &= \{\alpha_s x(1-x)[x + s^*(1-2x)] \\ &\quad - \alpha_u x + \alpha_v(1-x)\}(2N)^{-1} + o(N^{-1}) \\ var[\delta x | x] &= x(1-x)(2N)^{-1} + o(N^{-1}) \\ E[|\delta x|^3] &= o(N^{-1}). \end{aligned} \quad (3.1.34)$$

Supondo que uma unidade de tempo no processo de difusão corresponda a $2N$ gerações na cadeia de Markov chegamos a

$$\delta t = (2N)^{-1} \quad (3.1.35)$$

$$a(x) = \alpha_s x(1-x)[x + h(1-2x)] - \alpha_u x + \alpha_v(1-x) \quad (3.1.36)$$

$$b(x) = x(1-x), \quad (3.1.37)$$

que são coeficientes da equação de difusão.

A equação de difusão em si fica

$$\frac{\partial z}{\partial t}(x; p, t) = a(p) \frac{\partial z}{\partial p}(x; p, t) + \frac{1}{2} b(p) \frac{\partial^2 z}{\partial p^2}(x; p, t) \quad (3.1.38)$$

sendo $z(x; p, t)$ a densidade de probabilidade de transição da condição inicial p no instante t para a condição final x . Em geral, busca-se uma solução na forma de expansão em auto-funções

$$z(x; p, t) = \sum_{i=1}^{+\infty} c_i(x, p) e^{(-\lambda_i t)}.$$

Maiores detalhes sobre a técnica de aproximação por difusão e sua motivação podem ser apreendidos diretamente de FELLER [11], EWENS [9], [18], MARUYAMA [22].

3.2 Modelo de Moran

O **modelo de Moran** data de 1958 [23]. Ele permite obter expressões explícitas para muitas medidas de interesse evolucionário, mas aplica-se apenas a populações haplóides.

Hipóteses 3.2 (Modelo de Moran).

M1 - O número de indivíduos presentes na população é fixo e igual a $2N$.

M2 - A análise é feita sobre um locus gênico;

M3 - Dois alelos A e a são observados para esse locus.

M4 - A fertilidade e a sobrevivência são independentes do genótipo.

M5 - Não há mutação.

M6 - Os indivíduos são haplóides.

Em $t = 1, 2, 3, \dots$ um indivíduo é escolhido aleatoriamente para se reproduzir. Após a reprodução um indivíduo é escolhido para morrer (exceto o novo indivíduo). Considere $2N$ indivíduos haplóides, cada um sendo A ou a . Seja X_n o número de indivíduos do tipo A no tempo n . No tempo $n + 1$ haverá:

- $X_n + 1$ indivíduos A , se A for escolhido para reproduzir, e a para morrer.
- $X_n - 1$ indivíduos A , se a for escolhido para reproduzir, e A para morrer.
- X_n indivíduos A , se A for escolhido para reproduzir e morrer, ou se a for escolhido para reproduzir, mas também morrer.

Modelo de Moran

Temos

$$Prob(X_{n+1} = X_n + 1 | X_n = i) = \frac{i}{2N} \frac{2N - i}{2N} \quad (3.2.39)$$

$$Prob(X_{n+1} = X_n - 1 | X_n = i) = \frac{i}{2N} \frac{2N - i}{2N} \quad (3.2.40)$$

$$Prob(X_{n+1} = X_n | X_n = i) = \frac{i}{2N} \frac{i}{2N} + \frac{2N - i}{2N} \frac{2N - i}{2N} \quad (3.2.41)$$

Note que

$$E[X_{n+1} | X_n = i] = i \quad (3.2.42)$$

Assim $E[X_{n+1}] = E[E[X_{n+1} | X_n]] = E[X_n]$ e portanto $E[X_n] = E[X_0]$.
Pode-se obter facilmente também [29],

$$E[X_{n+1}^2 | X_n = i] = 2 \frac{i(2N - i)}{(2N)^2}. \quad (3.2.43)$$

Exercício 3.9. *Demonstre as expressões (3.2.42) e (3.2.43).*

3.2.1 Taxa de decaimento de heterozigotos

Assim como no modelo de Wright-Fisher, eventualmente um dos genes irá se fixar na população e tem-se

$$P_i[X_\tau = 2N] = \frac{i}{2N},$$

onde $\tau = \min\{n : X_n = 0 \text{ ou } X_n = 2N\}$. Isto é, a probabilidade do gene A_1 se fixar na população, sob as hipóteses do modelo de Moran, é igual a sua freqüência inicial. A demonstração segue o raciocínio da apresentada na seção 3.1.1.

Calculando o valor esperado da heterozigosidade, temos

$$E[H_{n+1} | X_n = i] = 2 \left(\frac{i}{2N} \right) \left(1 - \frac{i}{2N} \right) \left(1 - \frac{2}{N^2} \right) \quad (3.2.44)$$

e portanto

$$E[H_{n+1}] = E[H_n] \left(1 - \frac{2}{N^2} \right) \quad (3.2.45)$$

Recursivamente obtemos

$$E[H_n] = E[H_0] \left(1 - \frac{2}{N^2} \right)^n \quad (3.2.46)$$

Para n grande, $E[H_n]$ tende a

$$E[H_n] = E[H_0]e^{\frac{-2n}{(2N)^2}} \quad (3.2.47)$$

A fim de comparar com a deriva no modelo de Wright-Fisher, devemos definir a *geração* no modelo de Moran de forma a ser igual a $2N$ eventos de reprodução e morte. De fato, o tempo de vida T de um indivíduo pode ser medido da seguinte forma. Como um indivíduo tem probabilidade $\frac{1}{2N}$ de ser escolhido para morrer num passo de tempo, então a probabilidade de que ele sobreviva por $n - 1$ períodos e só morra no próximo é

$$P(T = n) = \left(1 - \frac{1}{2N}\right)^{n-1} \frac{1}{2N} \quad (3.2.48)$$

ou seja, T segue uma distribuição geométrica. Temos

$$E[T] = 2N \quad (3.2.49)$$

que é a esperança de vida de um indivíduo, podendo ser interpretada como duração de uma geração. Tomando $\tilde{n} = \frac{n}{2N}$ e substituindo na equação (3.2.47) vem que

$$E[H_{\tilde{n}}] = E[H_0]e^{\frac{-2\tilde{n}}{2N}} \quad (3.2.50)$$

Assim, com definições equivalentes de geração, a taxa de deriva genética é duas vezes mais rápida no modelo de Moran do que no de Wright-Fisher, dado pela equação (3.1.14). A diferença se deve à estrutura de reprodução que é distinta em cada um dos modelos [29].

3.2.2 Mutaçãõ

No modelo de Moran, para incluir mutaçãõ, vamos considerar que quando um indivíduo do tipo a for escolhido para reproduzir, com probabilidade v ele se transforma em A . Se um indivíduo A for escolhido para se reproduzir, então com probabilidade u ele via a . Assim se houver i indivíduos do tipo A na população, a probabilidade de escolher A para a próxima geração é

$$p_i = \frac{i}{2N}(1 - u) + \frac{2N - i}{2N}v. \quad (3.2.51)$$

As probabilidades de transiçãõ no modelo de Moran ficam

$$\begin{aligned} p_{i,i+1} &= \text{Prob}(X_{n+1} = X_n + 1 | X_n = i) = \frac{2N-i}{2N}p_i \\ p_{i,i-1} &= \text{Prob}(X_{n+1} = X_n - 1 | X_n = i) = \frac{i}{2N}(1 - p_i) \\ p_{i,i} &= \text{Prob}(X_{n+1} = X_n | X_n = i) = (1 - p_i)\frac{2N-i}{2N} + \frac{i}{2N}p_i \end{aligned} \quad (3.2.52)$$

Modelo de Moran

Se procurarmos uma distribuição de probabilidades sobre os possíveis estados nos quais o processo de Moran pode se encontrar, e que não se altere com o passar do tempo, devemos procurar $\pi = \{\pi_0, \pi_1, \dots, \pi_{2N}\}$ onde π_i é a probabilidade de haver i genes A , que satisfaz ainda,

$$\pi_i = \pi_{i-1}p_{i-1,i} + \pi_i p_{i,i} + \pi_{i+1}p_{i+1,i} \quad (3.2.53)$$

Isto é, a probabilidade de i genes A permanecerem, depende da transição de $i-1 \rightarrow i$ a partir do estado $i-1$, ou da transição de $i \rightarrow i$ a partir do estado i , ou da transição de $i+1 \rightarrow i$ a partir do estado $i+1$. Uma distribuição de probabilidades satisfazendo (3.2.53) é dita **distribuição estacionária**.

A equação (3.2.53) fornece $2N + 1$ equações, sendo que para $i = 0$ e $i = 2N$ temos

$$\pi_0 = \pi_0 p_{0,0} + \pi_1 p_{1,0} \quad (3.2.54)$$

$$\pi_{2N} = \pi_{2N-1} p_{2N-1,2N} + \pi_{2N} p_{2N,2N} \quad (3.2.55)$$

o que implica

$$\pi_0 = \pi_0(1 - v) + \pi_1 \frac{1}{2N}(1 - p_1) \quad (3.2.56)$$

$$\pi_{2N} = \pi_{2N-1} \frac{1}{2N} p_{2N-1} + \pi_{2N}(1 - u) \quad (3.2.57)$$

Exercício 3.10. Prove que $\pi_0 = \frac{\pi_1 p_{1,0}}{p_{0,1}}$, $\pi_1 = \frac{\pi_2 p_{2,1}}{p_{1,2}}$, $\pi_2 = \frac{\pi_3 p_{3,2}}{p_{2,3}}$, etc. De forma geral:

$$\pi_i = \frac{\pi_{i-1} p_{i-1,i}}{p_{i,i-1}} \quad (3.2.58)$$

para $i = 0, \dots, 2N - 1$. Note que $p_{i,i} = 1 - p_{i,i+1} - p_{i,i-1}$.

Da equação (3.2.58) obtemos de imediato que

$$\pi_i = \frac{p_{i-1,i} p_{i-2,i-1} p_{i-3,i-2} \dots p_{k+1,k+2} p_{k,k+1}}{p_{i,i-1} p_{i-1,i-2} p_{i-2,i-3} \dots p_{k+2,k+1} p_{k+1,k}} \pi_k = \prod_{j=k+1}^i \frac{p_{j-1,j}}{p_{j,j-1}} \quad (3.2.59)$$

Uma aproximação para a distribuição estacionária (3.2.59) é dada da seguinte forma. Suponha que N seja grande e sejam $q = 2Nv$ e $r = 2Nu$. A distribuição estacionária do processo de Moran, quando reescalada para o intervalo $[0, 1]$ se aproxima da distribuição beta (q, r) de densidade

$$f(x) = c_{q,r} x^{q-1} (1-x)^{r-1} \quad (3.2.60)$$

onde $c_{q,r}$ é uma constante de normalização tal que $\int_0^1 f(x) dx = 1$. Para a dedução ver DURRETT [5].

3.2.3 Aproximação por difusão

Da mesma forma que o modelo de Wright-Fisher, o modelo de Moran possui uma aproximação por difusão, que no caso sem seleção, é dada por,

$$\frac{\partial u}{\partial t}(x; p, t) = \frac{1}{4}p(1-p) \frac{\partial^2 z}{\partial p^2}(x; p, t) \quad (3.2.61)$$

onde $u(x; p, t)$ é a densidade de probabilidade de transição da condição inicial p no instante t para a condição final x .

Notas

Os modelos de Wright-Fisher e de Moran são o tema do capítulo 2 de DURRETT [5], cuja linguagem é acessível para biólogos e matemáticos. Idéias intuitivas sobre deriva genética aleatória podem ser vistas em GILLESPIE [14]. A aproximação por difusão do modelo de Wright-Fisher encontra-se descrito em detalhes no livro de EWENS [9]. Para difusões veja-se também KARLIN & TAYLOR [18] e MARUYAMA [22]. Uma introdução para teoria dos coalescentes é WAKELEY [29].

Bibliografia

- [1] BRITTON, N., “Essential Mathematical Biology”, Springer, London, 2003.
- [2] BÜRGER, R., “The Mathematical Theory of Selection, Recombination, and Mutation”, John Wiley & Sons, Chichester, 2000.
- [3] DA SILVA, Telles Timóteo, “Contribuições à Genética Populacional via Processos de Fleming-Viot”, Tese de Doutorado, LNCC, Petrópolis, RJ, 2006.
- [4] DAWSON, D. A., “Measure-valued Markov Processes”. In: Hennequin, P. L. (ed), *École d’Été de Probabilités de Saint-Flour XXI. Lecture Notes in Math. 1541*, Berlin, Springer-Verlag, pp.1-260, 1993.
- [5] DURRETT, R. “Probability Models for DNA Evolution”, Springer, New York, 2002.
- [6] EDWARDS, A. W. F., “The Fundamental Theorem of Natural Selection”, *Theoretical Population Biology*, v. 61, pp. 335-337, 2002.
- [7] ETHERIDGE, A., *An Introduction to Superprocesses*. 1st. Providence, American Mathematical Society, 2000.
- [8] ETHIER, S. N. & KURTZ, T. G., “Fleming-Viot processes in population genetics”, *SIAM J. Control and Optimization*, v. 31, n. 2, pp. 345-386, 1993.
- [9] EWENS, W. J., “Mathematical Population Genetics”, Springer, New York, 1979.
- [10] EWENS, W. J., “The changing role of population genetics theory”, *Lecture Notes in Biomathematics*, v. 100, pp. 186-197, 1994.

- [11] FELLER, W., “Diffusion Processes in Genetics”. In: Neyman, J. (ed), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, pp. 227-246, 1951.
- [12] FLEMING, Wendell H., “Diffusion processes in population biology”, *Supp. Adv. Appl. Prob.*, v. 7, pp. 100-105, 1975.
- [13] FLEMING, W. & VIOT, M., “Some measure-valued Markov processes in population genetics theory”, *Indiana Univ. Math. J.*, v. 28, n. 5, pp. 817-843, 1979.
- [14] GILLESPIE, J. H., “Population Genetics: a concise guide”, 2ed., The John Hopkins University Press, Baltimore, London, 2004.
- [15] GRAUR, D.& LI, W., “Fundamentals of Molecular Evolution”, 2ed., Sinauer Associates, Sunderland, Massachusetts, 2000.
- [16] HIRABA, S., “Jump-Type Fleming-Viot Processes”, *Adv. Appl. Prob.*, v. 32, pp. 140-158, 2000.
- [17] HOFBAUER, Josef, SIGMUND, Karl, *Evolutionary Games and Population Dynamics*. 1st. Cambridge, Cambridge University Press, 1998.
- [18] KARLIN, S. & TAYLOR, H., “A Second Course in Stochastic Processes”, Academic Press, New York, 1981.
- [19] KINGMAN, J.F.C., “The Coalescent”, *Stochastic Processes an their Applications*, v. 13, pp. 235-248, 1982.
- [20] KURTZ, Thomas G., *Aproximation of Population Processes*. 1st. Philadelphia, SIAM, 1981.
- [21] LEHNINGUER, A. L., NELSON, D. L. & COX, M. M., “Princípios de Bioquímica”, 4ed., Sarvier, São Paulo, 2006.
- [22] MARUYAMA, T., “Stochastic Problems in Genetics”, Springer, Berlin, 1977.
- [23] MORAN, P. A. P., “Random processes in genetics”, *Proc. Camb. Phil. Soc.*, v. 54, pp. 60-71, 1958.
- [24] OHTA, T. & KIMURA, M., “A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population”, *Genet. Res. Camb*, v. 22, pp. 201-204, 1973.

Bibliografia

- [25] PAGE, R. & HOLMES, E., “Molecular Evolution: a phylogenetic approach”, Blackwell Science, 2001.
- [26] WATSON, BAKER, BELL, GANN, LEVINE, LORICK, “Biologia molecular do gene”, 4ed., Artmed, Porto Alegre, 2006.
- [27] WAKELEY, John, “Recent Trends in Population Genetics: More Data! More Math! Simple Models?”, *Journal of Heredity*, v. 95, pp. 397-405, 2004.
- [28] WAKELEY, J., “The Limits of Theoretical Population Genetics”, *Genetics*, v. 169, pp. 1-7, 2005.
- [29] WAKELEY, J., “Coalescent Theory: an introduction”, Roberts & Company Publishers, Greenwood Village, Colorado, 2009.