

MINICURSO – COLÓQUIO DE MATEMÁTICA DA REGIÃO NORTE
2014

Comitê Científico

Flávia Morgana de O. Jacinto (UFAM) - Coordenadora

Hugo Alex Carneiro Diniz (UFOPA)

Jorge Herbert Soares de Lira (UFC)

Marcelo Miranda Viana da Silva (IMPA-SBM)

Renato de Azevedo Tribuzy (UFAM)

Rodrigo Bissacot Proença (USP)

Rúbia Gonçalves Nascimento (UFPA)

Esta é mais uma publicação da Sociedade Brasileira de Matemática para os minicursos ministrados nos Colóquios.

Os autores que submeterem propostas de minicursos devem estar cientes de que o texto deve ser preparado em **Latex (compatível com o MikTeX versão 2.7)**, as figuras em **eps**.

O texto deve ser redigido de forma clara e recomendamos a inclusão de **exercícios** para a verificação de aprendizagem, ao final de cada capítulo.

Veja outras publicações da SBM, na livraria virtual que se encontra na
página

<http://www.loja.sbm.org.br/>

Sociedade Brasileira de Matemática

2014

Aplicações biológicas de cadeias de Markov

Armando G. M. Neves
aneves@mat.ufmg.br

Departamento de Matemática
Instituto de Ciências Exatas
Universidade Federal de Minas Gerais

Sociedade Brasileira de Matemática

Rio de Janeiro - RJ, Brasil
2014

Coordenação Editorial:

Flávia Morgana de O. Jacinto

Editora: SBM

Impresso na Gráfica:

Capa: ? ? ?

Patrocínio: Superintendência da Zona Franca de Manaus (SUFRAMA)

Copyright ©2014 by Autores
Direitos reservados, 2014 pela SBM.

**Catálogo elaborado pela Biblioteca ???
Bibliotecária: ????**

Neves, Armando G. M.

Aplicações biológicas de cadeias de Markov – Rio de Janeiro, RJ :
SBM, 2014, ?? p., 20.5 cm - (Minicurso Colóquio CO 2014; v. ??)

ISBN ????-????

1. Palavra chave 1 2. Palavra chave 2 3. Palavra chave 3
I. Autor1, Nome 1. , II. Autor2, Nome 2 III. Título. IV. Série

CDD - 51

*Ao professor Armando Gil de Almeida Neves,
que estudou Medicina, mas não foi médico,
a não ser em emergências,
que foi professor de Bioquímica, mas me ensinou a ser
cientista,
que falava também sobre Matemática e Física,
que foi meu pai
e me ensinou também o que os pais ensinam aos filhos.*

Agradecimentos

Agradeço à queridíssima Márcia e aos filhos preciosos Francisco e Eduardo, que vivem comigo, me ouvem falar de Matemática, Genética, Neandertais e Física, me aturam e me inspiram. Amo-os mais do que tudo.

Agradeço ao Maurizio Serva, coautor no artigo sobre a extinção dos Neandertais, por ter-me chamado a atenção para algo que eu conhecia, mas ainda não tinha apreciado: a deriva genética.

Agradeço à SBM e à organização do Colóquio da Região Norte pela oportunidade de lecionar um minicurso. Agradeço aos prováveis alunos do minicurso, que irão tornar mais humanas estas palavras e fórmulas.

A vida entrava como uma senhora despótica: ele não a conhecia, e mesmo assim ela entrava em seu corpo, em sua mente; entrava como versos, como uma inspiração. E o significado dessa palavra pela primeira vez revelou-se a ele em toda sua plenitude.

...

Anotar, publicar: tudo isso é mera vaidade. Tudo que nasce por interesse não é o melhor. O melhor é o que não foi anotado, que foi composto e desapareceu, que se dissipou sem deixar rastro; e somente a alegria criadora, que ele percebe e que não se pode misturar com nada, mostra que o poema foi criado, que o belo foi criado.

Do conto Xerez, de Varlam Chalámov (1907-1982), tradução de Nivaldo dos Santos.

Conteúdo

Prefácio	13
1 Probabilidade, processos estocásticos e cadeias de Markov	17
1.1 Breve introdução à teoria das probabilidades	17
1.2 Cadeias de Markov	24
1.3 Dinâmica das cadeias de Markov	30
1.4 Alguns resultados sobre matrizes	35
1.5 Voltando à dinâmica das cadeias de Markov	40
1.6 Classificação de estados em cadeias de Markov	43
2 O processo de Moran generalizado	51
2.1 Processos de nascimento e morte	52
2.2 Probabilidade de fixação	55
2.3 Tempo médio de fixação	60
2.4 Populações pequenas e deriva genética	65
3 Um modelo para a extinção dos Neandertais	69
3.1 A história natural dos hominídeos	69
3.2 Modelo para a mistura genética de duas subpopulações	72
3.3 O processo de Wright-Fisher	75
3.4 Aplicando o modelo de mistura	81
A Distribuições binomial e geométrica	87
A.1 Distribuição binomial	87
A.2 Distribuição geométrica	88

Prefácio

Em Física, se nenhuma força resultante age em um corpo, sua aceleração é nula. O corpo pode portanto se mover, mas sua velocidade é constante. Uma consequência é que, na ausência de força resultante, não pode mudar o sentido da velocidade.

Os biólogos evolucionistas usam a analogia da Física e falam das “forças” que direcionam a mudança das frequências com que os genes aparecem em uma população. Darwin descobriu que a seleção natural é a força mais intensa que age sobre a Evolução: os mais aptos sobrevivem, ou, dito de outra maneira, genes que de alguma forma produzem mais descendentes tendem a ter sua frequência aumentada na população.

Mais tarde, por volta dos anos 1930, Sewall Wright e Ronald Fisher descobriram algo paradoxal: mesmo na ausência da “força” de seleção natural e de outras “forças” como mudanças do tamanho da população e migrações, as frequências genéticas em uma população podem se mover. Mas, ao contrário da inércia da Física, em que o movimento sem força é em um sentido único, o movimento das frequências genéticas na ausência de forças pode mudar de sentido várias vezes e levar à extinção genes mesmo quando não conferem menor aptidão. Este fenômeno ficou conhecido como *deriva genética*, ou genetic drift em inglês. Embora não iremos explorar muito mais analogias com a Física, a deriva genética tem grande semelhança matemática com um outro movimento errático e aparentemente sem força, que intrigou os físicos até ser desvendado por Einstein: o *movimento Browniano*.

Mas, voltando à história, Ronald Fisher achava que o papel da deriva genética era pouco importante em Biologia e este ponto de vista foi o dominante por muito tempo. Até que, em 1968, Motoo Kimura lançou a sua *teoria neutra da evolução molecular*, que afirma basicamente que a maioria das mudanças genéticas em uma população é resultado da deriva genética.

Este livro é basicamente sobre a teoria matemática por trás da deriva genética. Por incompetência do autor, não iremos entrar na polêmica exis-

tente até hoje entre os que defendem a maior ou menor importância da deriva genética como “força” que impulsiona a evolução biológica. Por outro lado, o objetivo é usar ferramentas matemáticas não muito sofisticadas: Cálculo Diferencial e Integral, Álgebra Linear e noções de Probabilidade, tudo isto ao nível de estudantes com uma boa maturidade e que tenham completado (ou quase) os dois primeiros anos de um curso na área de Ciências Exatas. O germe deste livro foi uma parte de uma disciplina de introdução à Biologia Matemática que lecionei na UFMG no 1º semestre deste ano para alguns estudantes dos cursos de graduação em Matemática e Matemática Computacional.

O capítulo 1 é a parte mais teórica, contendo passo a passo aquilo que achamos necessário para poder culminar com o Teorema 1.7, que é o que irá “produzir” a deriva genética nos modelos de Genética de Populações. A parte o Teorema de Perron-Frobenius e alguns resultados correlatos, enunciados sem demonstração [? ? ?], todos os demais resultados ou estão aqui demonstrados, ou a demonstração segue linhas bem delineadas no texto. O assunto *Cadeias de Markov* é bastante conhecido e existe vastíssima literatura a respeito [? ? ?], mas a maioria dos textos é dirigida a leitores mais avançados e apresenta uma teoria bastante geral, que servirá também para cadeias com espaços de estados infinitos, e não costuma ir além de uma pequena menção ao Teorema de Perron-Frobenius. Embora falemos um pouco também dos outros casos, nosso foco principal são as cadeias com espaços de estados *finitos*. Daí a importância assumida pela Teoria de Perron-Frobenius, sobre a qual se baseiam alguns dos resultados mais importantes do capítulo 1.

Já no capítulo 2 saímos do geral e particularizamos, falando sobre os chamados *processos de nascimento e morte*, que incluem como caso especial o modelo de Moran. Não pudemos evitar falar no problema da ruína do jogador, que não é uma das aplicações biológicas que tínhamos em mente, mas estava tão ao alcance da mão... Aprendemos também como calcular probabilidades e tempos de fixação em cadeias de Markov, que são fundamentais para as aplicações biológicas.

No capítulo 3 introduzimos uma aplicação da deriva genética como explicação para a extinção dos Neandertais. Para isto é necessário introduzir também o processo de Wright-Fisher que, mais do que o processo de Moran, desempenha o papel de “principal” modelo matemático em Genética de Populações. Os resultados desse capítulo foram publicados em artigo do autor em colaboração com Maurizio Serva, da Università dell’Aquila, Itália.

O livro termina com dois apêndices onde, em benefício de quem não tenha porventura cursado uma disciplina de Probabilidade, definimos duas das mais importantes distribuições de probabilidade discretas – a binomial

e a geométrica – e calculamos para elas a média e a variância. Ambas possuem papel central em nossa exposição. A distribuição binomial aparece no processo de Wright-Fisher e a geométrica na prova do Teorema 1.7.

E para terminar este prefácio situo o leitor no ambiente do conto Xerez, de Varlam Chalámov, citado na epígrafe. Esse narra a agonia do poeta russo Óssip Mandelstam (1891-1938). Preso em uma Gulag, padecendo a fome, nos seus derradeiros instantes ele pensava na forma de versos, que obviamente não ficaram para os pósteros.

Mais do que publicar este livro, a alegria de tê-lo pensado é que ficará em meu coração.

Belo Horizonte, 29 de agosto de 2014.

Armando G. M. Neves

Capítulo 1

Probabilidade, processos estocásticos e cadeias de Markov

1.1 Breve introdução à teoria das probabilidades

A Teoria de Probabilidades é um grande ramo dentro da Matemática. Não pretendemos aqui nada de muito profundo a respeito. Aqueles que o desejarem poderão procurar em outros livros, por exemplo [? ? ?], onde o assunto é tratado com mais profundidade, em qualquer nível desde o elementar até o muito avançado. Vamos nos ater aqui ao mínimo que julgamos adequado para chegar aos objetivos a que nos propomos neste texto.

Seja Ω um conjunto, que chamaremos *espaço amostral*. Os subconjuntos de Ω serão chamados de eventos. A teoria de probabilidades começa quando atribuímos a cada evento $A \subset \Omega$ um número $P(A)$ que será chamado a *probabilidade de A*. Naturalmente não chamaremos probabilidade a qualquer função que atribua valores numéricos aos subconjuntos de Ω . A definição a seguir restringe de maneira adequada o conjunto de funções admissíveis.

Definição 1.1. *Seja $\mathcal{P}(\Omega)$ o conjunto de todos os subconjuntos de Ω . Uma função $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$ será dita uma probabilidade se:*

$$(i) P(A) \geq 0, \quad \forall A \in \mathcal{P}(\Omega)$$

(ii) $P(\emptyset) = 0$

(iii) $P(\Omega) = 1$

(iv) Se A_1, A_2, \dots é uma família enumerável de eventos mutuamente disjuntos, ou seja $A_i \cap A_j = \emptyset$ se $i \neq j$, então $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$.

Antes que sejamos acusados de imprecisão, cabe observar que a definição acima não está estritamente correta. Em muitos exemplos importantes é impossível atribuir uma probabilidade a *todo* subconjunto de Ω . Nesses casos é necessário restringir o domínio da função P a um subconjunto (em geral bastante grande para conter todos os eventos relevantes) de $\mathcal{P}(\Omega)$. Tal subconjunto deve ser uma σ -álgebra. Para maiores detalhes o leitor deve consultar um livro sobre Medida e Integração, por exemplo [? ?]. Em particular, uma função definida em uma σ -álgebra de conjuntos que cumpra os requisitos (i), (ii) e (iv) da definição acima é chamada uma *medida*. Em todos os casos que aparecerão neste texto será possível definir uma probabilidade para todos os subconjuntos do espaço amostral e não precisaremos nos preocupar com questões envolvendo σ -álgebras.

Exemplo 1.1. Lançamento de dois dados

Considere a situação comum em que temos dois dados distintos, por exemplo um primeiro dado vermelho e um segundo dado amarelo. Vamos supor ainda que nenhum dos dados é viciado, ou seja, cada um deles cairá com a mesma frequência em cada uma das 6 faces. Queremos atribuir probabilidades para os possíveis resultados do experimento de lançar os dois dados.

Neste caso, definimos $R = \{1, 2, 3, 4, 5, 6\}$ como sendo o conjunto dos possíveis resultados dos lançamentos de um único dado e nosso espaço amostral será $\Omega = R \times R$, conjunto com 36 elementos.

Cada elemento de Ω será um par ordenado (d_1, d_2) , onde $d_1 \in R$ denota o resultado do lançamento do primeiro dado e $d_2 \in R$ será o resultado do lançamento do segundo dado.

A probabilidade de qualquer $A \in \mathcal{P}(\Omega)$ será definida por

$$P(A) = \frac{\text{número de elementos de } A}{36}.$$

O leitor deve verificar que essa função satisfaz a todos os requisitos da Definição 1.1. O fator $\frac{1}{36}$ foi colocado de modo a satisfazer (iii) e o fato de todos os conjuntos com um único elemento terem a mesma probabilidade é o que nos diz que os dados não são viciados, de forma que qualquer dos 36 possíveis resultados para o lançamento dos dois dados é igualmente provável.

Com esta definição, podemos por exemplo calcular a probabilidade de que o primeiro dado caia com o número 6. Seja $D_{16} = \{(d_1, d_2) \in \Omega; d_1 = 6\}$, Claramente, $D_{16} = \{(6, 1), (6, 2), \dots, (6, 6)\}$ tem 6 elementos. Então

$$P(D_{16}) = \frac{6}{36} = \frac{1}{6}.$$

Definição 1.2. *Sejam A_1 e A_2 eventos em Ω . Se $P(A_2) \neq 0$, podemos definir a probabilidade condicional de A_1 dado A_2 por*

$$P(A_1|A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)}. \quad (1.1.1)$$

O sentido da definição acima é o seguinte: suponha que sabemos que o evento A_2 ocorreu; então $P(A_1|A_2)$ é a probabilidade de que ocorra A_1 entre todos os eventos em que A_2 também ocorreu. Para entender melhor o conceito acima, vamos introduzir um segundo conceito e, a seguir, mais um exemplo:

Definição 1.3. *Dois eventos A_1 e A_2 são ditos independentes se*

$$P(A_1 \cap A_2) = P(A_1)P(A_2). \quad (1.1.2)$$

A observação a seguir explica de forma intuitiva o significado da definição acima. Se A_1 e A_2 são independentes, então, por (1.1.2),

$$P(A_1|A_2) = \frac{P(A_1)P(A_2)}{P(A_2)} = P(A_1).$$

Ou seja, se A_1 e A_2 são independentes, então a ocorrência de A_2 não modifica nem para mais, nem para menos, a probabilidade de ocorrência de A_1 .

Exemplo 1.2. *Novamente o lançamento de dois dados*

Considere novamente os dados do exemplo 1.1 e o evento D_{16} ali introduzido. Seja $D_{21} = \{(d_1, d_2) \in \Omega; d_2 = 1\}$ o evento de o segundo dado cair com resultado 1. Com a definição de probabilidade dada no exemplo 1.1 temos claramente $P(D_{21}) = \frac{1}{6}$.

Intuitivamente sabemos que o fato de primeiro dado ter caído com resultado 1 não deve influenciar de forma alguma o resultado do segundo dado. De fato, a probabilidade de o segundo dado cair em 1, sabendo que o primeiro caiu em 6 é

$$P(D_{21}|D_{16}) = \frac{P(D_{21} \cap D_{16})}{P(D_{16})} = \frac{1}{6},$$

pois $D_{21} \cap D_{16} = \{(6, 1)\}$ possui probabilidade $\frac{1}{36}$, já que é conjunto com um único elemento. O resultado $P(D_{21}|D_{16}) = \frac{1}{6}$ é o mesmo que teríamos obtido intuitivamente quando pensamos que o resultado do segundo dado não é afetado pelo do primeiro dado. Essa intuição é corretamente captada pelo conceito de independência da Definição 1.3. Observe que, de fato, $P(D_{21} \cap D_{16}) = P(D_{21})P(D_{16})$ e portanto, pela Definição 1.3, esses dois eventos são independentes.

Definição 1.4. *Uma variável aleatória (real) é uma função $X : \Omega \rightarrow \mathbb{R}$. Se a imagem $X(\Omega)$ de uma variável aleatória é um conjunto enumerável (ou finito), dizemos que a variável aleatória é discreta. Caso contrário, dizemos que é uma variável aleatória contínua. A imagem $X(\Omega)$ será chamada de espaço de estados de X .*

Neste texto só trabalharemos com variáveis aleatórias discretas. Só por completeza é bom deixar o leitor avisado de que definição acima, embora adequada aos fins do presente texto, não é a mais correta no caso geral. Para torná-la correta basta que o leitor adicione o adjetivo *mensurável* à função, mas isto mais uma vez nos leva a questões como σ -álgebras e à teoria de Medida e Integração.

Aqui interrompemos brevemente a sequência de definições para introduzir uma notação conveniente. Se X é uma variável aleatória e $x \in \mathbb{R}$, denotamos por $\{X = x\}$ o evento de a variável aleatória X tomar o valor x , ou seja,

$$\{X = x\} \equiv \{\omega \in \Omega; X(\omega) = x\} = X^{-1}(\{x\}).$$

Mais geralmente, se $A \subset \mathbb{R}$, então $\{X \in A\}$ será o evento $\{\omega \in \Omega; X(\omega) \in A\} = X^{-1}(A)$.

Podemos naturalmente associar probabilidades aos conjuntos de valores assumidos por uma variável aleatória X . Esta associação é feita por

$$P(X \in A) \equiv P(\{X \in A\}). \quad (1.1.3)$$

Em particular, temos:

Definição 1.5. *Se X é uma variável aleatória discreta, então a função $f : \mathbb{R} \rightarrow \mathbb{R}$ tal que*

$$f(x) = P(X = x)$$

é dita a função (massa) de probabilidade de X .

Exemplo 1.3. *Lançamento de N moedas*

Suponha que sejam jogadas N moedas independentes e que para cada uma delas a probabilidade de sair cara (diremos que o resultado da moeda é

1 se cair em cara) seja igual a p . Uma moeda sair em coroa (diremos que o resultado foi 0) naturalmente terá uma probabilidade $q = 1 - p$. Caso $p = \frac{1}{2}$ todas as moedas são honestas, mas não consideraremos necessariamente que isto acontece.

O espaço amostral é $\Omega = \{(\omega_1, \dots, \omega_N) \in \mathbb{R}^N; \omega_i = 0 \text{ ou } \omega_i = 1, i = 1, 2, \dots, N\}$.

Temos várias possibilidades interessantes de variáveis aleatórias neste exemplo. Por exemplo, $X_i(\omega) = \omega_i$ é a variável aleatória que dá o resultado da i -ésima moeda no lançamento ω . Um outro exemplo que vamos explorar novamente mais adiante é

$$C(\omega) = \sum_{i=1}^N \omega_i,$$

que calcula o número de moedas que caiu em cara em um lançamento.

O conceito de independência de eventos pode ser estendido a variáveis aleatórias:

Definição 1.6. *Sejam X_1, \dots, X_n variáveis aleatórias sobre o mesmo espaço amostral. Diremos que as variáveis X_1, \dots, X_n são independentes se os eventos $\{X_1 \in A_1\}, \{X_2 \in A_2\}, \dots, \{X_n \in A_n\}$ forem independentes para qualquer escolha de conjuntos $A_1, A_2, \dots, A_n \subset \mathbb{R}$.*

Continuamos agora com uma definição dentre as mais importantes e bem conhecidas de forma intuitiva:

Definição 1.7. *Seja X variável aleatória discreta com espaço de estados $\{a_1, a_2, \dots\}$. Seja $f(a_i) = P(X = a_i)$ a função de probabilidade de X . A média (ou valor esperado) de X é definida por*

$$\mu_X \equiv \sum_{i=1}^{\infty} a_i f(a_i). \quad (1.1.4)$$

Observe que a soma acima pode não convergir se o espaço de estados de X for infinito.

Antes de continuar, é hora de introduzir uma notação bastante usada: se X é variável aleatória discreta com espaço de estados $\{a_1, a_2, \dots\}$ e função de probabilidade f e g uma função cujo domínio contenha o espaço de estados de X então $\mathbb{E}(g(X))$ é definido como

$$\mathbb{E}(g(X)) = \sum_{i=1}^{\infty} g(a_i) f(a_i), \quad (1.1.5)$$

de modo que a média de X é simplesmente $\mu_X = \mathbb{E}(X)$. O operador \mathbb{E} é chamado de *operador esperança*.

Observe que uma maneira de medir o quanto uma variável aleatória se afasta de sua média poderia ser calculando a esperança da diferença entre a variável e sua média, ou seja,

$$\begin{aligned}\mathbb{E}(X - \mu_X) &= \sum_{i=1}^{\infty} (a_i - \mu_X) f(a_i) = \sum_{i=1}^{\infty} a_i f(a_i) - \sum_{i=1}^{\infty} \mu_X f(a_i) \\ &= \mu_X - \mu_X \sum_{i=1}^{\infty} f(a_i) = 0.\end{aligned}$$

pois $\sum_{i=1}^{\infty} f(a_i) = P(\Omega) = 1$.

Como $\mathbb{E}(X - \mu_X)$ é sempre nula, a maneira padrão de se medir quanto uma variável aleatória se afasta de sua média é a *variância*, definida a seguir:

Definição 1.8. *A variância de uma variável aleatória discreta X com espaço de estados $\{a_1, a_2, \dots\}$ e função de probabilidade f é*

$$\text{Var}_X \equiv \mathbb{E}((X - \mu_X)^2). \quad (1.1.6)$$

Não é difícil provar que $\text{Var}_X \geq 0$ e que também pode ser calculada como

$$\text{Var}_X = \mathbb{E}(X^2) - \mu_X^2. \quad (1.1.7)$$

Exemplo 1.4. *Novamente o lançamento de N moedas*

Continuamos agora o Exemplo 1.3. Primeiramente precisamos atribuir probabilidades aos eventos. Naturalmente, atribuímos probabilidade nula ao evento vazio, de modo a satisfazer (ii) na Definição 1.1. Como cada cara possui probabilidade p e cada coroa probabilidade q , então a probabilidade de uma jogada com c caras e $N - c$ coroas é $p^c q^{N-c}$. Mais formalmente, lembrando que $\omega_i = 1$, se a i -ésima moeda caiu em cara, $\omega_i = 0$ se caiu em coroa, a probabilidade de um evento correspondente a um único lançamento das N moedas é

$$P(\{(\omega_1, \dots, \omega_N)\}) = p^{\sum_{i=1}^N \omega_i} q^{N - \sum_{i=1}^N \omega_i}. \quad (1.1.8)$$

A definição acima pode ser estendida aos demais eventos de Ω simplesmente por aditividade: uma vez que sabemos a probabilidade de todos os eventos unitários, ou seja eventos com um único elemento, basta somar as probabilidades de cada um de seus elementos para obter a probabilidade de um evento com mais elementos.

Para verificar que faz sentido a definição acima de probabilidade para os eventos, é necessário verificar que os requerimentos da Definição 1.1 foram todos satisfeitos. É trivial verificar (i), e (ii) já foi naturalmente incluído na construção. Também (iv) é verdadeiro por causa da aditividade que usamos para definir a probabilidade de eventos com mais de um elemento. Falta mostrar que (iii) também vale. Para isto, observe que o número n_c de eventos unitários em que c das N moedas caíram em cara, $c = 0, 1, 2, \dots, N$, é

$$n_c = \binom{N}{c} \equiv \frac{N!}{c!(N-c)!}. \quad (1.1.9)$$

Podemos então calcular $P(\Omega)$ somando sobre todos os possíveis valores de c :

$$\begin{aligned} P(\Omega) &= \sum_{c=0}^N \sum_{\omega \in \Omega; \sum_{i=1}^N \omega_i = c} P(\omega) \\ &= \sum_{c=0}^N \binom{N}{c} p^c q^{N-c} = (p+q)^N = 1. \end{aligned}$$

Na última linha usamos (1.1.9), (1.1.8) e a fórmula do binômio de Newton.

A média e a variância da variável aleatória C definida no exemplo 1.3 serão de grande importância neste texto (e em diversas outras instâncias). Intuitivamente, se temos N moedas independentes e cada uma cai em cara com probabilidade p , então o número médio de caras em um lançamento deve ser

$$\mu_C = Np. \quad (1.1.10)$$

A variância de C é dada por

$$Var_C = Npq. \quad (1.1.11)$$

Embora não saibamos uma maneira intuitiva de justificar essa fórmula como fizemos como (1.1.10), ambas são cuidadosamente deduzidas no Apêndice A. Para terminar o exemplo e deixar mais claro o significado do conceito de variância, vale a pena explorar aqui uma consequência importante de (1.1.11).

Se o número N de moedas lançadas é muito grande, é de se esperar que o número de caras em um lançamento seja muito próximo da média μ_C . Uma vez que a variância de uma variável aleatória X nos dá a esperança do quadrado da diferença de X para a sua média μ_X , então $\sqrt{Var_X}$ é uma maneira de se medir os valores absolutos típicos das diferenças entre X e sua média. No caso particular da variável C , temos $\sqrt{Var_C} = \sqrt{pq}\sqrt{N}$ e

essa diferença não tende a 0 quando $N \rightarrow \infty$, pois também a média μ_C diverge quando $N \rightarrow \infty$. No entanto, quando comparamos $\sqrt{\text{Var}_C}$ com μ_C vemos que, se $p \neq 0$, $\frac{\sqrt{\text{Var}_C}}{\mu_C} = \frac{\sqrt{q/p}}{\sqrt{N}}$ tende a 0 quando $N \rightarrow \infty$. Portanto, no limite $N \rightarrow \infty$, a variância torna-se desprezível quando comparada à média.

1.2 Cadeias de Markov

Revisando algumas fontes diferentes, por exemplo [? ? ? ?], encontramos algumas definições parecidas, mas não equivalentes, para o termo *cadeia de Markov*. A definição a seguir não é inteiramente consensual, mas será aquela suficientemente interessante para ser usada nestas notas.

Definição 1.9. *Uma sequência infinita de variáveis aleatórias $(X_n)_{n=0}^\infty$ é chamada uma cadeia de Markov se:*

1. *Todas as variáveis X_n possuem o mesmo espaço de estados E .*
2. *O espaço de estados E é enumerável (ou finito).*
3. *Se i_0, i_1, \dots, i_n são elementos arbitrários de E , então*

$$\begin{aligned} P(X_n = i_n \mid X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \dots, X_1 = i_1, X_0 = i_0) \\ = P(X_n = i_n \mid X_{n-1} = i_{n-1}). \end{aligned} \quad (1.2.12)$$

4. *Para quaisquer $x, y \in E$, $n \in \mathbb{N}$,*

$$P(X_n = x \mid X_{n-1} = y) = P(X_1 = x \mid X_0 = y). \quad (1.2.13)$$

A ideia por trás dessa definição é que o valor da variável aleatória X_n represente o valor no tempo n de alguma grandeza em um sistema que evolua de forma aleatória. Em geral acontece que X_n não é independente de X_{n-1} .

Uma sequência de variáveis aleatórias que obedece somente aos requisitos 1 e 2 acima seria chamada um *processo estocástico discreto a tempo discreto*. A equação (1.2.12) que distingue as cadeias de Markov de outros processos estocásticos é chamada a *propriedade de Markov* e (1.2.13) é usualmente chamada de *condição de homogeneidade*. A propriedade de Markov significa que a regra probabilística para mudar do estado i_{n-1} no tempo $n-1$ para o estado i_n no tempo n não depende dos tempos anteriores $0, 1, \dots, n-2$. Muitos autores descrevem essa propriedade dizendo que o sistema perde a memória do que aconteceu antes e só considera o estado no tempo $n-1$.

para decidir o estado no tempo n . A condição (1.2.13) estabelece que as probabilidades de mudança de um estado para outro são independentes do tempo.

Dois exemplos dentre modelos matemáticos interessantes para fenômenos biológicos irão esclarecer melhor processos estocásticos em geral e cadeias de Markov em particular.

No primeiro exemplo veremos um processo estocástico que *não* é uma cadeia de Markov.

Exemplo 1.5. *O modelo SIR estocástico*

Considere uma população de tamanho constante – N indivíduos – sujeita a uma doença contagiosa. Após serem infectados, os indivíduos adquirem a doença e se curam com certeza após um certo tempo, mas adquirem imunidade permanente, de modo a que nunca tornam a se infectar. Uma doença com essas características é, por exemplo, a rubéola.

A população pode então ser naturalmente dividida em 3 classes de indivíduos: os susceptíveis S são aqueles que nunca tiveram a doença, os infectados I são aqueles que no momento estão doentes e podem transmitir a doença para os susceptíveis e os recuperados R são aqueles que já tiveram a doença, se curaram, não mais a transmitem e tampouco voltam a adoecer.

Na versão determinística desse modelo, bem conhecida [? ?], os números $S(t), I(t), R(t)$ de susceptíveis, infectados e recuperados no instante t são dados pelas soluções do seguinte sistema de equações diferenciais ordinárias:

$$\begin{cases} S'(t) &= -\frac{\beta}{N}S(t)I(t) \\ I'(t) &= \frac{\beta}{N}S(t)I(t) - \gamma I(t) \\ R'(t) &= \gamma I(t) \end{cases}, \quad (1.2.14)$$

onde β e γ são parâmetros positivos que caracterizam o sistema. Para maiores detalhes a respeito, o leitor deve consultar algum livro sobre modelos matemáticos em Biologia, por exemplo, os citados acima. O parâmetro β pode ser pensado como a “força” da infecção, englobando tanto o número de contatos entre os indivíduos da população, quanto a efetividade desses contatos em transmitir de um para outro a doença. O parâmetro γ é basicamente o inverso do tempo necessário para que um indivíduo infectado se recupere.

Uma versão estocástica bem simples para esse modelo pode ser obtida supondo que o tempo é discreto e a unidade de tempo, digamos uma semana, é o tempo que um indivíduo infectado leva para se recuperar. A dinâmica da versão estocástica é obtida da seguinte maneira: a cada instante de tempo n cada um dos infectados sorteia de maneira independente z indivíduos na população, que são aqueles com quem ele teve contato naquela unidade de

tempo. Caso um indivíduo contactado por um infectado seja susceptível, então tornar-se-á infectado no tempo $n + 1$; caso contrário, nada acontece com ele. Além do mais, todos os infectados no instante n passarão à classe dos removidos no instante $n + 1$.

Supondo que a doença é muito infecciosa, ou seja, que cada contato entre um infectado e um susceptível levem a uma infecção do susceptível, a razão $\frac{\beta}{N}$ na versão determinística do modelo corresponde ao parâmetro z da versão estocástica.

Na Figura 1.1 mostramos o número de infectados como função do tempo na versão determinística e em uma realização dos sorteios na versão estocástica.

Figura 1.1: Número de infectados como função do tempo nos modelos SIR determinístico e estocástico. Em ambos os casos, a população total é $N = 400$. A linha contínua é obtida através da solução numérica do sistema (1.2.14) com $\beta = \frac{1}{200}$, $\gamma = 1$, e condição inicial de 1 infectado, nenhum recuperado e 399 susceptíveis. Os pontos se referem a uma realização dos sorteios da versão estocástica com $z = 2$ e mesma condição inicial de 1 infectado e 399 susceptíveis.

Tipicamente, enquanto o número total de infectados e recuperados é pequeno o suficiente, o número de infectados cresce rapidamente com o tempo. No entanto, em um certo tempo os infectados e recuperados já são tão abundantes na população que o encontro entre um infectado e um susceptível passa a ser raro em confronto com o encontro entre um infectado e um indivíduo que não possa adquirir a doença. A partir daí o número de infectados passa a decrescer com o tempo até a extinção da epidemia. Isto

acontece tanto na versão determinística, quanto na versão estocástica.

No caso estocástico, o modelo SIR é um processo estocástico discreto a tempo discreto $(I_n)_{n=0}^{\infty}$, onde as variáveis aleatórias são os números de infectados I_n nos instantes de tempo discreto n . O fato de a curva típica de número de infectados como função do tempo possuir um máximo mostra claramente que *não vale* a propriedade Markov (1.2.12). Uma maneira de ver isto é pensar que I_n depende naturalmente de I_{n-1} , mas também de I_k para instantes $k \leq n-2$. De fato, os infectados em instantes $k \leq n-2$ estarão recuperados no instante $n-1$ e o número de recuperados claramente afeta o número de susceptíveis e portanto o resultado do sorteio que definirá I_n .

Vamos agora a um exemplo de cadeia de Markov descrevendo uma outra situação biologicamente interessante, e que vamos explorar mais a fundo neste texto.

Exemplo 1.6. *O processo de Moran neutro*

Vamos descrever – em sua forma mais simples – aquilo que em Genética de Populações é conhecido como o Processo de Moran *neutro*. Este processo é um *modelo* útil para entender quantitativa e qualitativamente as mudanças na composição genética de uma população à medida que o tempo passa, alguns indivíduos morrendo e outros se reproduzindo.

Considere uma população com tamanho fixo de N indivíduos, como no modelo SIR. Suponha que nessa população existam dois tipos de indivíduos, que chamaremos indivíduos de tipo A e de tipo B. Por exemplo, podemos pensar que A e B sejam duas formas alternativas de um mesmo gene (os geneticistas diriam que A e B são dois alelos para um mesmo locus gênico). Suponha que o tempo seja discreto, como no exemplo anterior, e que no instante n , $n = 0, 1, 2, \dots$, sejam realizados dentre todos os indivíduos da população dois sorteios *independentes e uniformes*: sortearemos um indivíduo que irá se reproduzir (estamos portanto supondo que a reprodução é assexuada, o que é uma simplificação útil) e um indivíduo (que pode ser o mesmo escolhido para se reproduzir) que irá morrer. Por sorteios uniformes queremos dizer que em cada um dos sorteios cada indivíduo da população possui a mesma probabilidade de ser sorteado. A composição da população no instante $n+1$ será então igual à composição da população no tempo n com exceção da seguinte mudança: um indivíduo do mesmo tipo do indivíduo sorteado para se reproduzir irá substituir o indivíduo sorteado para morrer. Estamos portanto supondo que o tipo de um indivíduo é sempre o mesmo tipo do seu ancestral. Em termos genéticos, esta suposição significa que estamos desprezando a possibilidade de mutação no tipo de um indivíduo.

Com as regras estipuladas no parágrafo anterior o número N de indivíduos na população será constante, pois a cada instante de tempo somente um indivíduo irá morrer e somente um indivíduo irá contribuir com um descendente. Seja X_n o número de indivíduos de tipo A no instante n , $X_n \in \{0, 1, 2, \dots, N\}$. Cada X_n , $n = 0, 1, 2, \dots$ é uma variável aleatória com espaço de estados $E = \{0, 1, \dots, N\}$. O processo acima descrito é portanto um processo estocástico a tempo discreto com espaço de estados discreto.

Pela regra de reprodução e morte, vemos que se $X_n = k$, $k \in E$, $k \neq 0$ e $k \neq N$, então X_{n+1} pode assumir somente 3 valores:

1. Caso sejam sorteados um indivíduo A para se reproduzir e um indivíduo A para morrer, então $X_{n+1} = k$, pois o indivíduo A que morreu será substituído por outro A descendente do que se reproduziu. Este caso ocorre com probabilidade $(\frac{k}{N})^2$, pois em cada um dos dois sorteios a probabilidade de sortear um A é $\frac{k}{N}$.
2. Analogamente, também acontecerá que $X_{n+1} = k$ caso sejam sorteados indivíduos de tipo B para se reproduzir e para morrer. Isto acontecerá com probabilidade $(\frac{N-k}{N})^2$, pois a probabilidade de sortear um B é $\frac{N-k}{N}$ em cada um dos dois sorteios.
3. Caso seja sorteado um indivíduo de tipo A para se reproduzir e um indivíduo de tipo B para morrer, então teremos $X_{n+1} = k + 1$ pois um indivíduo de tipo A irá substituir um indivíduo morto de tipo B. A probabilidade de acontecer este caso é $\frac{k}{N} \frac{N-k}{N}$.
4. Finalmente, caso um indivíduo de tipo B seja sorteado para se reproduzir e um indivíduo de tipo A sorteado para morrer, então $X_{n+1} = k - 1$. A probabilidade deste caso também é $\frac{k}{N} \frac{N-k}{N}$.

Caso se tenha $X_n = 0$ ou $X_n = N$ então em um caso não haverá nenhum indivíduo de tipo A para ser possivelmente sorteado e no outro caso não haverá indivíduos de tipo B. Caso aconteça um dos dois casos então certamente $X_{n+1} = X_n$ e $X_m = X_n$ para todo $m \geq n$. Diremos que no caso $X_m = 0 \quad \forall m \geq n$ houve *fixação* do tipo B na população. No outro caso, diremos que houve *fixação* do tipo A.

Uma observação importante é que dada a população no tempo n e escolhidos arbitrariamente dois indivíduos nessa população, a probabilidade de que o primeiro indivíduo seja sorteado para se reproduzir é exatamente a mesma probabilidade de que o segundo indivíduo seja o sorteado para a reprodução, *independentemente do tipo do primeiro e do segundo indivíduos*.

Além do mais, o mesmo vale para a probabilidade dos mesmos indivíduos serem sorteados para morrer.

Em outras palavras, o tipo A ou B de um indivíduo é, neste exemplo, uma característica que não influencia nem positiva, nem negativamente, as suas probabilidades de sobrevivência ou de reprodução. Do ponto de vista da seleção natural o tipo A ou B é uma característica *neutra*. No capítulo 2 iremos considerar também variantes deste exemplo em que o tipo A ou B pode tornar um indivíduo mais ou menos apto para a sobrevivência ou para a reprodução.

Olhando novamente as 4 possibilidades de sorteios enumeradas acima e as consequências que cada uma delas tem para o valor de X_{n+1} , observamos que o estado X_{n+1} só depende do estado X_n , não dependendo do número de indivíduos de tipo A em instantes anteriores a n . Isto é exatamente a propriedade de Markov (1.2.12). Além do mais, a regra probabilística para mudança de um estado para outro depende só do estado e não do instante n em que ocorre a mudança. Isto é a condição de homogeneidade (1.2.13). Com isto, provamos que o processo estocástico $(X_n)_{n=0}^{\infty}$ é uma cadeia de Markov. Provamos ainda que a regra explícita para mudança de estado é

$$P(X_{n+1} = j | X_n = k) = \begin{cases} \frac{k}{N} \frac{N-k}{N}, & \text{se } j = k + 1 \\ \frac{k}{N} \frac{N-k}{N}, & \text{se } j = k - 1 \\ 1 - 2\frac{k}{N} \frac{N-k}{N}, & \text{se } j = k \\ 0, & \text{se } |j - k| \geq 2 \end{cases}, \quad (1.2.15)$$

onde é importante notar que os casos particulares $k = 0$ e $k = N$ já estão incluídos na fórmula acima, pois nesses casos haverá probabilidade 1 de que $j = k$ e 0 para qualquer outro valor de j .

Antes de encerrar, mostramos na Figura 1.2 os gráficos dos valores de X_n em função de n para 5 realizações da dinâmica do processo de Moran neutro com população $N = 50$. Com isto queremos dizer que programamos o computador para realizar os sorteios de reprodução e morte por diversas unidades de tempo seguidas e fizemos isto 5 vezes. Em cada realização tomamos sempre a mesma condição inicial $X_0 = 30$, mas como a evolução no processo de Moran é aleatória, naturalmente obtivemos resultados diferentes para as variáveis X_n em cada uma das realizações.

Um fato absolutamente notável na Figura 1.2 é que, apesar da neutralidade do processo, um dos dois tipos de indivíduos acabou extinto da população em cada uma das realizações mostradas: houve fixação em todas as realizações. Será que o autor está mostrando realizações atípicas do processo? Será que houve um erro do computador?

A neutralidade é manifesta na equação (1.2.15) se repararmos que a probabilidade de aumento de X_n de uma unidade é exatamente igual à

Figura 1.2: Simulações computacionais do processo de Moran neutro com população $N = 50$. Em todos os casos o número inicial de indivíduos de tipo A era $X_0 = 30$. Em 3 das simulações houve fixação do tipo A e nas outras 2 do tipo B. Em todos os casos, antes de 2000 unidades de tempo houve fixação de um dos dois tipos de indivíduos.

probabilidade de decréscimo de uma unidade. Uma primeira intuição é que em um processo como este poderá haver alguma flutuação de X_n em torno de X_0 , mas que a probabilidade de um afastamento grande de X_n do valor inicial X_0 deve ser pequena e, portanto, a fixação deveria ser muito improvável para valores grandes da população total N , pelo menos se X_0 não é próximo de nenhum dos dois extremos 0 e N .

Um dos resultados mais interessantes que iremos provar nestas notas é que a intuição acima está *errada* e que o comportamento ilustrado na Figura 1.2 é absolutamente típico: a fixação ocorre com probabilidade 1. Apesar da neutralidade, ou seja, aqui não há ação da seleção natural, há uma “força” que leva um dos dois tipos à extinção. Esse é o fenômeno da *deriva genética*.

1.3 Dinâmica das cadeias de Markov

Em um processo estocástico com a propriedade de Markov (1.2.12) podemos definir uma *matriz de transição* como sendo uma matriz T_n (possivelmente infinita) onde o elemento da j -ésima linha e k -ésima coluna é a probabilidade de acontecer no tempo n a transição do estado k para o estado j . Se além do mais vale a homogeneidade (1.2.13), então T_n é independente

de n . Isto motiva a seguinte

Definição 1.10. *Seja X_n cadeia de Markov. A matriz de transição T da cadeia é a matriz (possivelmente infinita) tal que*

$$t_{jk} = P(X_1 = j | X_0 = k). \quad (1.3.16)$$

Aqui devemos observar que a matriz de transição é por enquanto simplesmente uma maneira conveniente de tabular as probabilidades que definem de maneira unívoca uma cadeia de Markov. Obviamente o lado direito de (1.3.16) poderia talvez mais naturalmente ser definido como o elemento t_{kj} em vez de t_{jk} e a maior parte dos livros que tratam cadeias de Markov assim o faz. Vamos porém seguir aqui, por razões que ficarão claras logo, a convenção contrária, também adotada por outros livros. Ou seja, t_{jk} será para nós a probabilidade da transição do estado k para o estado j .

Uma consequência simples da definição da matriz de transição é que cada um de seus elementos é um número real no intervalo $[0, 1]$, pois é uma probabilidade, e além do mais a soma de cada coluna da matriz é sempre igual a 1. De fato, os elementos da k -ésima coluna de T são as probabilidades de transição do estado k para cada um dos outros estados. Seja E o espaço de estados da cadeia. Como os eventos $\{X_{n+1} = i\}$, $i \in E$ são mutuamente disjuntos e $\bigcup_{i \in E} \{X_{n+1} = i\} = \Omega$, então os requerimentos (iii) e (iv) da Definição 1.1 implicam que

$$\sum_{i \in E} t_{ik} = 1.$$

Matrizes que obedecem às características acima são ditas *estocásticas por colunas*.

Definição 1.11. *Uma matriz M $n \times n$, onde também podemos ter $n = \infty$, é dita estocástica por colunas se seus elementos m_{ij} são números reais não-negativos e para cada j tem-se*

$$\sum_{i=1}^n m_{ij} = 1. \quad (1.3.17)$$

Como já vimos, cadeias de Markov modelam situações em que o futuro é, em geral, incerto e queremos acompanhar ao longo dos instantes de tempo $n = 0, 1, 2, \dots$ os valores X_n de alguma grandeza do sistema. Mesmo que conheçamos com certeza o valor inicial X_0 da grandeza, em geral só nos é dado conhecer as probabilidades de cada um dos possíveis valores para X_1 . Podemos generalizar essa situação e pensar que também X_0 seja conhecido

somente como uma distribuição de probabilidades de ocupação de cada um dos possíveis estados. Portanto, em geral, uma cadeia de Markov no instante n , $n = 0, 1, 2, \dots$ será descrita por um vetor v_n com tantas componentes quantos forem os estados das variáveis aleatórias X_k e tal que sua i -ésima componente é

$$(v_n)_i = P(X_n = i). \quad (1.3.18)$$

Obviamente, tem que valer

$$\sum_{i=1}^{\infty} (v_n)_i = 1. \quad (1.3.19)$$

Definição 1.12. Um vetor v é dito um vetor de probabilidade se todas suas componentes são números em $[0, 1]$ e se a soma de suas componentes é igual a 1.

O resultado a seguir mostra a utilidade de se tabular como uma matriz as probabilidades de transição que definem a cadeia.

Proposição 1.1. Seja $(X_n)_{n=0}^{\infty}$ cadeia de Markov com espaço de estados E e matriz de transição T . Se $v_n \in \mathbb{R}^E$ é um vetor de probabilidade cujas componentes $(v_n)_i$ indicam para cada $i \in E$ a probabilidade de que X_n seja igual a i , então

(i) $T v_n$ também é um vetor de probabilidade.

(ii) Se definirmos $v_{n+1} = T v_n$, então $P(X_{n+1} = j) = (v_{n+1})_j$.

Em outras palavras, a matriz de transição atua como operador linear em \mathbb{R}^E de evolução temporal da cadeia de Markov.

Demonstração: (i) Vamos supor por generalidade que E é infinito. Caso E seja finito a prova é a mesma.

$$(v_{n+1})_j = \sum_{k=1}^{\infty} t_{jk}(v_n)_k.$$

Como os elementos t_{jk} e as componentes $(v_n)_k$ são todos números não-negativos, claramente as componentes de v_{n+1} também o serão.

Falta somente provar que a soma das componentes de v_{n+1} também é igual a 1.

$$\sum_{j=1}^{\infty} (v_{n+1})_j = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} t_{jk}(v_n)_k = \sum_{k=1}^{\infty} (v_n)_k \sum_{j=1}^{\infty} t_{jk},$$

onde na última passagem fizemos somente uma troca de ordem de soma, trivial no caso em que E é finito, mas que o leitor deve tentar justificar caso E seja infinito. Usando a propriedade de que T é estocástica por colunas e também (1.3.19) obtemos o desejado.

(ii) Observe que

$$\{X_{n+1} = j\} = \bigcup_{k \in E} (\{X_{n+1} = j\} \cap \{X_n = k\}) ,$$

pois $\bigcup_{k \in E} \{X_n = k\} = \Omega$. Como os $\{X_n = k\}$, $k \in E$ são mutuamente disjuntos, usando (iv) na Definição 1.1 temos

$$\begin{aligned} (v_{n+1})_j = P(X_{n+1} = j) &= \sum_{k \in E} P(\{X_{n+1} = j\} \cap \{X_n = k\}) \\ &= \sum_{k \in E} P(X_{n+1} = j | X_n = k) P(X_n = k) \\ &= \sum_{k \in E} t_{jk} (v_n)_k , \end{aligned}$$

onde usamos ainda a definição de probabilidade condicional, ver (1.1.1). ■

Corolário 1.1. *Se definirmos $T^{(n)}$ como a matriz de transição de n passos de uma cadeia de Markov, ou seja, a matriz cujos elementos são*

$$(T^{(n)})_{jk} = P(X_n = j | X_0 = k) ,$$

então

$$T^{(n)} = T^n , \tag{1.3.20}$$

ou seja, a matriz de transição a n passos é a n -ésima potência da matriz de transição.

Em muitos casos a evolução de sistemas biológicos determinísticos é dada por um sistema de equações diferenciais, ver por exemplo (1.2.14) que descreve a evolução de uma epidemia. Em muitos casos, as equações diferenciais possuem soluções de equilíbrio. Também frequentemente, soluções diferentes das de equilíbrio podem convergir assintoticamente para as soluções de equilíbrio. Por exemplo, pode-se mostrar que todas as soluções de (1.2.14) com condições iniciais positivas para S , I e R convergem quando $t \rightarrow \infty$ para soluções de equilíbrio em que o número de infectados é nulo, como mostra a Figura 1.1.

É possível mostrar que cadeias de Markov *com espaço de estados finito* sempre possuem distribuições de probabilidade de equilíbrio, ou seja, vetores

de probabilidade v que são invariantes pela evolução temporal da cadeia. Também veremos que sob certas condições as distribuições de equilíbrio de cadeias de Markov são distribuições assintóticas, ou seja, a cadeia tende à distribuição de equilíbrio no limite de tempo infinito.

Várias das afirmativas acima também continuam valendo para cadeias com espaço de estados infinitos. Para evitar certas questões de convergência e algumas delicadezas relativas a espaços de dimensão infinita, preferiremos neste texto apresentar de agora em diante somente resultados para cadeias de Markov finitas. Esperamos que o leitor, como nós, possa achá-los suficientemente ricos e interessantes.

Iniciamos aqui a sequência de resultados que nos levarão a entender a existência de distribuições de equilíbrio e a convergência para elas. A primeira coisa a se observar é que um vetor de probabilidade w é estacionário pela evolução de uma cadeia de Markov se acontecer que ao termos $v_n = w$, então acontecerá também $v_{n+1} = w$ (e portanto também $v_m = w$ para todo $m > n$). Mas, pela Proposição 1.1, $v_{n+1} = Tv_n$. Portanto, observe que um vetor de probabilidade w é estacionário por uma cadeia de Markov somente se for um autovetor da matriz de probabilidade T com autovalor 1. O resultado a seguir mostra que para cadeias finitas temos sempre *candidatos* a distribuições de equilíbrio.

Teorema 1.1. *Se T é matriz $n \times n$ estocástica por colunas, então 1 é autovalor de T .*

Demonstração: Seja T^t a transposta da matriz T . Seja $\mathbf{1} \in \mathbb{R}^n$ o vetor com todas as componentes iguais a 1. Do fato de que a soma de cada uma das linhas de T^t é 1 resulta que

$$T^t \mathbf{1} = \mathbf{1}.$$

Portanto, 1 é autovalor de T^t . Como os polinômios característicos de uma matriz e de sua transposta são idênticos, então também são idênticos seus autovalores. Isto prova que 1 é autovalor de T . ■

Conforme dizíamos, o teorema acima ainda não prova a existência de distribuições estacionárias para cadeias de Markov finitas. De fato, não sabemos se um autovetor com autovalor 1 da matriz de transição T pode ser interpretado como um vetor de probabilidade. Pois, apesar de podermos normalizá-lo de maneira a que a soma das componentes seja igual a 1, não sabemos ainda se podemos tomar todas as componentes desse autovetor como não-negativas.

Para entender melhor essa questão, assim como outras que aparecerão, na próxima seção falaremos sobre alguns resultados a respeito de matrizes não-negativas.

1.4 Alguns resultados sobre matrizes

A teoria de Perron-Frobenius trata de autovalores de matrizes cujos elementos são todos não-negativos. Perron primeiramente provou um resultado sobre matrizes cujos elementos são todos positivos e Frobenius enxergou o conceito de irredutibilidade e conseguiu estender o resultado de Perron. O assunto é tratado em profundidade em alguns livros como [? ? ?], mas aqui vamos nos ater somente às definições e aos enunciados dos teoremas principais.

Definição 1.13. *Seja \mathbb{P} o conjunto dos reais não negativos, $\mathbb{P} = [0, \infty)$. Uma matriz A $n \times n$ é dita não-negativa se todos os elementos a_{ij} da matriz forem elementos de \mathbb{P} .*

Podemos associar às matrizes não-negativas digrafos (ou grafos dirigidos). Sem querer também enunciar de maneira abstrata o conceito de digrafo, podemos nos contentar aqui com a ideia de que um digrafo é um “desenho” onde *arestas*, ou seja, linhas *orientadas*, ligam *vértices*. Observe que as arestas de digrafos são orientadas, ou seja, são representadas com setas, indo de um vértice de partida a um vértice de chegada.

Definição 1.14. *O digrafo de adjacências de uma matriz $n \times n$ é o digrafo onde os vértices são os números $1, 2, \dots, n$ e existe uma aresta partindo do vértice j e chegando ao vértice i se e somente se $a_{ij} \neq 0$.*

Um exemplo deve deixar mais claro o conceito.

Exemplo 1.7. *Matriz de transição e digrafo de adjacências do processo de Moran*

De acordo com o Exemplo 1.6, o lado direito de (1.2.15) é o elemento t_{jk} da matriz de transição T da cadeia de Markov que descreve o processo de Moran neutro. O digrafo de adjacências de T , no caso $N = 10$, por exemplo, é apresentado na Figura 1.3. As linhas orientadas possuem aqui a interpretação de que somente as transições com probabilidade positiva entre um estado de partida k e um estado de chegada j estão assinaladas com arestas.

Definição 1.15. *Um caminho em um digrafo partindo do vértice j e chegando ao vértice i é um conjunto de arestas do grafo que, possivelmente passando por arestas intermediárias, saia do vértice j e chegue até o vértice i . O comprimento de um caminho que liga j a i é o número de arestas no caminho.*

Figura 1.3: Digrafo de adjacências para a matriz de transição do processo de Moran neutro com $N = 10$. Observe que os vértices de 1 a 9 estão ligados a si mesmos e a ambos os vértices vizinhos. Dos vértices 0 e 10 somente partem ligações para eles mesmos.

Definição 1.16. *Um digrafo é dito fortemente conexo se para quaisquer dois vértices j e i do digrafo existir um caminho no digrafo que parta de j e chegue até i .*

Exemplo 1.8. *Digrafo fortemente conexo*

O digrafo da Figura 1.4 é fortemente conexo. Observe que há caminhos que ligam qualquer vértice a qualquer outro. Por exemplo, o caminho $1 \rightarrow 2, 2 \rightarrow 4, 4 \rightarrow 3$ parte do vértice 1 e chega ao vértice 3.

Por outro lado, o digrafo da Figura 1.3, que representa as adjacências da matriz de transição do processo de Moran neutro, não é fortemente conexo, pois não existem caminhos que liguem o vértice 0 a qualquer outro vértice. Observe que, ao contrário, existem caminhos que ligam qualquer outro vértice ao vértice 0, mas isto não é suficiente para garantir que o digrafo é fortemente conexo. Da mesma forma, o vértice 10 também não está ligado por nenhum caminho aos demais vértices.

Conforme veremos adiante, os resultados mais poderosos da teoria de Perron-Frobenius dizem respeito a matrizes *irredutíveis*. O conceito de irredutibilidade será definido abaixo em termos do digrafo de adjacências da matriz correspondente. É possível definir, *de maneira equivalente*, uma matriz irredutível de modo puramente algébrico. Vamos preferir aqui a definição abaixo, que tem um sabor mais geométrico e mais adequado aos nossos propósitos. Para a definição algébrica e a prova da equivalência das duas definições referimos o leitor a [?].

Definição 1.17. *Uma matriz não-negativa é dita irredutível se o seu digrafo de adjacências for fortemente conexo.*

Antes de enunciar o teorema de Perron-Frobenius façamos um pequeno cálculo motivador, interessante por si só, e que não é restrito a matrizes não-negativas.

Seja A matriz $n \times n$ não-nula qualquer. Vamos supor por simplicidade que A seja diagonalizável sobre \mathbb{C} , mas o leitor que conheça a forma canônica

Figura 1.4: Exemplo de digrafo fortemente conexo.

de Jordan não terá dificuldade em verificar que a hipótese de diagonalizabilidade não é necessária aqui. Se A é diagonalizável, então existe uma base de \mathbb{C}^n formada por autovetores de A . Suponhamos que esses autovetores sejam v_1, \dots, v_n e que os tenhamos numerado em ordem decrescente do valor absoluto de seus autovalores. Ou seja, se $\lambda_1, \dots, \lambda_n$ são todos os autovalores, não necessariamente distintos, de A , sendo $\lambda_i \in \mathbb{C}$ o autovalor correspondente ao autovetor v_i , $Av_i = \lambda_i v_i$, então estamos supondo que

$$|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_n|.$$

Seja ainda $m \leq n$ o número de autovalores de A com valor absoluto máximo, ou seja,

$$|\lambda_i| = |\lambda_1| \quad \text{se } i = 1, 2, \dots, m.$$

Vamos agora tomar um vetor $w \in \mathbb{C}^n$ qualquer e calcular a ação sobre ele de A^k , onde $k \in \mathbb{N}$ é um número grande. Escrevendo w como combinação linear dos elementos da base de autovetores, $w = \sum_{i=1}^n c_i v_i$, temos

$$A^k w = A^k \sum_{i=1}^n c_i v_i = \sum_{i=1}^n c_i \lambda_i^k v_i.$$

Mas se A é não-nula, então $\lambda_1 \neq 0$. Podemos então colocar em evidência λ_1^k na última expressão e obteremos

$$A^k w = \lambda_1^k \sum_{i=1}^n c_i \left(\frac{\lambda_i}{\lambda_1} \right)^k v_i.$$

Nesta última expressão os números $\frac{\lambda_i}{\lambda_1}$ com $i = 1, 2, \dots, m$ têm todos módulo 1, enquanto para $i = m + 1, \dots, n$ todos têm módulo estritamente menor que 1. Estes últimos, quando elevados à potência k tenderão exponencialmente a 0 quando $k \rightarrow \infty$. Portanto se k é grande teremos

$$A^k w \approx \lambda_1^k \sum_{i=1}^m c_i \left(\frac{\lambda_i}{\lambda_1} \right)^k v_i .$$

A conclusão aqui é que, a menos do caso muito especial em que $c_1 = c_2 = \dots = c_m = 0$, se k é grande a ação de A^k tende a produzir um vetor aproximadamente na soma direta dos autoespaços de A com os autovalores de módulo máximo. Em particular, se $m = 1$, ou seja, A possui um único autovalor com módulo estritamente maior que todos os outros, $A^k w$ será aproximadamente um vetor no autoespaço de dimensão 1 correspondente ao autovalor λ_1 de módulo máximo.

Este cálculo motiva a definição seguinte:

Definição 1.18. *O maior autovalor em valor absoluto de uma matriz quadrada é chamado o autovalor dominante.*

Podemos agora enunciar o resultado principal sobre matrizes não-negativas. Para a prova e outros resultados correlatos, veja por exemplo [? ? ?].

Teorema 1.2 (Perron-Frobenius). *Se A é matriz não-negativa, então o autovalor dominante r de A é real e não-negativo. Além do mais, existe um autovetor v com autovalor r com todas as componentes não-negativas.*

Note que no Teorema 1.2 acima não incluímos a hipótese de que A é irredutível. O teorema a seguir sintetiza algumas propriedades importantes que obtemos como bônus quando essa hipótese adicional é incluída.

Teorema 1.3 (Perron-Frobenius para matrizes irredutíveis). *Suponha que A é matriz $n \times n$ não-negativa e irredutível. Então:*

- (i) *O autovalor dominante r de A é estritamente positivo e é raiz simples do polinômio característico de A .*
- (ii) *O autoespaço de A relacionado ao autovalor dominante possui dimensão 1.*
- (iii) *Existe um autovetor v correspondente ao autovalor dominante r com todas as componentes estritamente positivas.*
- (iv) *Os únicos autovetores de A em \mathbb{F}^n são os múltiplos não-negativos de v .*

Com relação ao cálculo de $A^k w$ que fizemos há pouco, a hipótese de irredutibilidade de A ainda não é suficiente para garantir que $m = 1$. Ou seja, apesar de o autoespaço relacionado ao autovalor dominante ser de dimensão 1, não se pode excluir a possibilidade de que A possua algum outro autovalor com módulo igual ao do autovalor dominante. Uma condição que garante que $m = 1$ é a chamada *primitividade*, ver [?]. Sem querer entrar em detalhes a respeito, o resultado abaixo é suficiente – mas não necessário – para garantir a condição de primitividade:

Teorema 1.4. *Se A é matriz $n \times n$ não-negativa e irredutível e seu traço $\text{Tr } A \equiv \sum_{i=1}^n a_{ii}$ é positivo, então o autovalor dominante de A é o único autovalor de A com módulo máximo.*

Interrompemos a seção anterior tendo *quase* provado que toda cadeia de Markov finita possui pelo menos uma distribuição de probabilidade de equilíbrio. Já tínhamos mostrado que a matriz de transição de qualquer cadeia de Markov finita admite 1 como autovalor. Faltava mostrar que o autovetor correspondente a esse autovalor podia ser interpretado como um vetor de probabilidade. A teoria de Perron-Frobenius esboçada até aqui será usada para provar que todas as componentes desse autovalor podem ser tomadas como não-negativas e isso será resultado do Teorema 1.2, desde que consigamos argumentar que 1 é o autovalor dominante da matriz de transição. Para fazê-lo vamos lançar mão de um resultado que vale para matrizes quaisquer, mas que se encaixa como uma luva ao que queremos. Começamos com uma definição:

Definição 1.19. *Seja A uma matriz quadrada $n \times n$ de elementos complexos. O i -ésimo disco de Gershgorin de A é o disco fechado \mathcal{D}_i no plano complexo centrado no ponto a_{ii} com raio*

$$R_i = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|,$$

ou seja,

$$\mathcal{D}_i = \{z \in \mathbb{C}; |z - a_{ii}| \leq R_i\}.$$

O resultado é tão geral, interessante, fácil de provar e pouco conhecido que não hesitamos nada em prová-lo. Ei-lo:

Teorema 1.5 (Gershgorin). *Seja A matriz $n \times n$ complexa qualquer. Suponha que $v = (v_1, \dots, v_n) \in \mathbb{C}^n$ é um autovetor de A com autovalor $\lambda \in \mathbb{C}$ e que para um certo k tem-se $|v_k| \geq |v_i|$, $i = 1, 2, \dots, n$. Então $\lambda \in \mathcal{D}_k$.*

Em particular, todos os autovalores de A se encontram na união $\bigcup_{i=1}^n \mathcal{D}_i$ dos discos de Gershgorin de A .

Se \mathcal{D}_i^t denota o i -ésimo disco de Gershgorin da transposta A^t de A , então vale também que os autovalores de A estão em $\bigcup_{i=1}^n \mathcal{D}_i^t$.

Demonstração: Seja λ autovalor de A , seja v um autovetor relacionado. Suponha que a maior componente de v em valor absoluto seja a k -ésima. A k -ésima componente da equação $Av = \lambda v$ é

$$\sum_{j=1}^n a_{kj}v_j = \lambda v_k .$$

Podemos reescrevê-la como

$$(\lambda - a_{kk})v_k = \sum_{j \neq k} a_{kj}v_j .$$

Passando o valor absoluto em ambos os membros, aplicando a desigualdade triangular e usando que a k -ésima componente de v é a maior em módulo, temos

$$|\lambda - a_{kk}||v_k| \leq \sum_{j \neq k} |a_{kj}||v_j| \leq |v_k| \sum_{j \neq k} |a_{kj}| .$$

Cancelando $|v_k|$ em ambos os lados, obtemos finalmente $\lambda \in \mathcal{D}_k$.

Em geral não sabemos qual das componentes de um autovetor será a maior em módulo e portanto pode ser difícil saber quais dos discos de Gershgorin de A irão conter o autovalor correspondente. No entanto, como alguma das componentes será a maior, podemos afirmar com certeza que todos os autovalores de A estarão em $\bigcup_{i=1}^n \mathcal{D}_i$.

Do fato de que os autovalores de A e A^t são idênticos segue a afirmativa sobre os discos de Gershgorin de A^t . ■

1.5 Voltando à dinâmica das cadeias de Markov

Tendo terminado um interlúdio com resultados interessantes sobre matrizes, podemos agora voltar a um assunto que tínhamos interrompido. Como consequência do Teorema de Gershgorin temos agora

Teorema 1.6. *O autovalor dominante de uma matriz A estocástica por colunas é igual a 1.*

Demonstração: Os discos de Gershgorin de A^t todos possuem centro no eixo real entre 0 e 1, todos possuem o ponto 1 em sua fronteira e todos têm raio menor ou igual a 1. Essas propriedades são consequência direta do fato de A ser estocástica por colunas, ver a Definição 1.11. A Figura 1.5 ilustra esses discos para um exemplo de matriz estocástica por colunas.

Sejam R_i e a_{ii} respectivamente raio e centro do disco de Gershgorin \mathcal{D}_i^t de A^t e λ autovalor qualquer de A . Então, caso $\lambda \in \mathcal{D}_i^t$ temos

$$|\lambda| \leq |\lambda - a_{ii}| + |a_{ii}| \leq R_i + a_{ii} = 1.$$

Como essa conclusão vale independentemente de i e como 1 é autovalor de A , segue o teorema. ■

Figura 1.5: Ilustração dos discos de Gershgorin da transposta de uma matriz estocástica por colunas.

Corolário 1.2. *Toda cadeia de Markov com espaço de estados finito possui uma distribuição de probabilidade de equilíbrio. Caso a matriz de transição da cadeia seja irredutível, tal distribuição de equilíbrio é única e todas as suas componentes são estritamente positivas.*

Demonstração: A existência de uma distribuição de equilíbrio vem do fato de que um autovetor do autovalor dominante possui componentes não-negativas, Teorema 1.2. Obviamente, podemos multiplicar tal autovetor por uma constante apropriada de modo a que seja satisfeito o requerimento de que a soma das componentes de um vetor de probabilidade seja igual a 1. A unicidade no caso de matrizes irredutíveis e a positividade das componentes vêm de (ii), (iii) e (iv) no Teorema 1.3. ■

A existência de distribuições de equilíbrio, a unicidade dessa distribuição, caso a matriz de transição seja irredutível, e a convergência da dinâmica da cadeia para uma distribuição de equilíbrio são propriedades importantes para cadeias de Markov. Para cadeias com espaços de estado finitos podemos provar essas propriedades usando a teoria de Perron-Frobenius, que é o ponto de vista que usamos aqui. No caso de cadeias infinitas outras ferramentas são necessárias, há fenômenos que não ocorrem em cadeias finitas e mais hipóteses são necessárias para poder-se provar alguns resultados importantes. Algumas referências que também mencionam e provam resultados para cadeias infinitas são [? ? ?].

No caso de cadeias finitas e com matriz de transição irredutível já sabemos da existência de uma única distribuição de equilíbrio, mas não é possível provar que qualquer distribuição de probabilidade inicial da cadeia converge para a distribuição de equilíbrio sem adicionar mais uma hipótese. O problema é que com as hipóteses até agora enunciadas ainda é possível que a matriz de transição possua algum autovalor de módulo 1 além do autovalor dominante 1. Quando aparece mais que um autovalor da matriz de transição com módulo igual a 1, a cadeia de Markov finita é dita *periódica*. Para se provar a convergência para o estado de equilíbrio de qualquer distribuição inicial precisamos garantir que a cadeia de Markov é *aperiódica*.

Não queremos entrar em mais detalhes a respeito de cadeias periódicas, pois elas não nos interessam aqui. Há porém uma hipótese verificada facilmente em muitos exemplos interessantes e que é *suficiente* para garantir aperiodicidade: $\text{Tr } T > 0$, ver Teorema 1.4.

Uma cadeia de Markov finita, irredutível e aperiódica é chamada *ergódica*. O conceito de ergodicidade existe também para cadeias infinitas, mas a definição requer mais condições.

O resultado abaixo sintetiza a discussão acima sobre convergência de cadeias de Markov para a distribuição de equilíbrio:

Corolário 1.3. *Considere cadeia de Markov com n estados e matriz de transição T irredutível, com $\text{Tr } T > 0$. Então, para qualquer vetor de probabilidade $w \in \mathbb{R}^n$,*

$$\lim_{k \rightarrow \infty} T^k w = v ,$$

onde v é a distribuição de equilíbrio da cadeia de Markov. Em outras palavras, qualquer que seja a distribuição de probabilidade w para o estado inicial, a distribuição de probabilidade da cadeia convergirá para a distribuição de equilíbrio.

A prova do resultado acima é basicamente uma reprodução do argumento que esboçamos mais atrás para motivar a definição de autovalor dominante. Vamos repetir aqui o argumento por causa de uma pequena diferença.

Demonstração: Vamos supor que T é diagonalizável em \mathbb{C} para simplificar, mas aqueles que conhecem a forma canônica de Jordan poderão usá-la aqui no lugar da diagonalizabilidade. Seja $\mathcal{B} = \{v, w_2, \dots, w_n\}$ uma base de \mathbb{C}^n composta por autovetores de T . Lembramos que v , a distribuição de equilíbrio da cadeia de Markov, é o único dos elementos de \mathcal{B} com autovalor 1. A hipótese $\text{Tr } T > 0$ implica ainda, ver Teorema 1.4, que os autovalores λ_j , $j = 2, \dots, n$ dos demais elementos de \mathcal{B} possuem módulo estritamente menor que 1. Se $w = x_0$ é o vetor de probabilidade que descreve a cadeia no instante inicial, então a distribuição da cadeia no tempo n é $x_n = T^n x_0$, ver o Corolário 1.1. Escrevendo x_0 na base \mathcal{B} , $x_0 = c_1 v + \sum_{j=2}^n c_j w_j$, obtemos $x_n = c_1 v + \sum_{j=2}^n c_j \lambda_j^n w_j$. Como $|\lambda_j| < 1$, $j = 2, \dots, n$, então $x_n \rightarrow c_1 v$ quando $n \rightarrow \infty$. Como x_n é vetor de probabilidade para qualquer instante n , Proposição 1.1, então temos que ter $c_1 = 1$. ■

1.6 Classificação de estados em cadeias de Markov

Nesta seção vamos introduzir nomenclaturas para distinguir diversos tipos de estados em cadeias de Markov. Veremos como resultados interessantes decorrerão dessa classificação.

No que segue consideraremos uma cadeia de Markov $(X_n)_{n=0}^{\infty}$ com espaço de estados E não necessariamente finito e matriz de transição T . Por $t_{ji}^{(n)}$ denotaremos os elementos da matriz T^n . Em particular, lembrando o Corolário 1.1, $t_{ji}^{(n)}$ é a probabilidade de que uma cadeia de Markov que se encontrava no estado i no instante inicial esteja no estado j após n unidades de tempo.

Definição 1.20. O estado $j \in E$ é acessível do estado i , condição que será denotada $i \rightarrow j$, se existir algum $n \geq 0$ tal que $t_{ji}^{(n)} > 0$. Em outras palavras, a condição $i \rightarrow j$ significa que a probabilidade de a dinâmica da cadeia de Markov efetuar a transição de i para j é não-nula, desde que se possa esperar um número suficiente de unidades de tempo.

A Proposição abaixo não é difícil de provar, mas é necessário fazer antes uma relação entre os elementos da n -ésima potência de uma matriz e os caminhos de comprimento n no digrafo de adjacências da matriz. Esta relação e a prova da Proposição 1.2 serão deixados para os exercícios. Temos então

Proposição 1.2. *A relação $i \rightarrow j$ vale se e somente se existe algum caminho no digrafo de adjacências de T que parta de i e chegue a j .*

Definição 1.21. *Os estados $i, j \in E$ são comunicantes se $i \rightarrow j$ e $j \rightarrow i$. Neste caso, denotaremos a situação por $i \leftrightarrow j$.*

Usando a Proposição 1.2 pode-se provar facilmente o resultado abaixo. A prova fica como exercício.

Proposição 1.3. *A relação \leftrightarrow é uma relação de equivalência. Ou seja:*

(i) *Se $i \in E$, então $i \leftrightarrow i$.*

(ii) *Se $i, j \in E$ e $i \leftrightarrow j$, então $j \leftrightarrow i$.*

(iii) *Se $i, j, k \in E$ e valem $i \leftrightarrow j$ e $j \leftrightarrow k$, então também se tem $i \leftrightarrow k$.*

Sabe-se de forma geral que qualquer relação de equivalência entre elementos de um conjunto particiona o conjunto em *classes de equivalência*. No caso atual, a relação de comunicação entre estados de uma cadeia de Markov particiona o espaço de estados E em classes de comunicação. A classe de comunicação de um estado i é o conjunto de todos os estados j tais que a cadeia possa realizar com probabilidade positiva, desde que se espere tempo suficiente, tanto a transição de i para j , quanto de j para i .

Definição 1.22. *Um estado $i \in E$ é dito absorvente se $t_{ii} = 1$*

Lembrando que as matrizes de transição T são estocásticas por colunas, se o estado i for absorvente, então o elemento da diagonal na coluna i é igual a 1 e todos os demais elementos dessa coluna são nulos.

Seja e_i o vetor de probabilidade em \mathbb{R}^E em que a coordenada correspondente ao estado i é igual a 1 e todas as demais são nulas. É fácil de ver que se i é absorvente, então e_i é distribuição de equilíbrio da cadeia de Markov.

Os estados 0 e N do processo de Moran neutro são exemplos de estados absorventes.

Definição 1.23. *A probabilidade de primeiro retorno ao estado i em n unidades de tempo é*

$$f_i^{(n)} = P(X_n = i, X_m \neq i, \text{ para } m = 1, 2, \dots, n-1 | X_0 = i).$$

Em palavras, condicionado a que o estado inicial da cadeia seja $X_0 = i$, $f_i^{(n)}$ calcula a probabilidade de que $X_n = i$ e que não se tenha $X_m = i$ para nenhum outro instante entre 0 e n .

Claramente, $\sum_{n=1}^{\infty} f_i^{(n)}$ é a probabilidade total de que a cadeia, começando em i , retorne a i em algum instante. Portanto, obviamente

$$\sum_{n=1}^{\infty} f_i^{(n)} \leq 1.$$

Isto motiva as definições seguintes:

Definição 1.24. Um estado $i \in E$ será dito *recorrente* se

$$\sum_{n=1}^{\infty} f_i^{(n)} = 1.$$

Um estado $i \in E$ será dito *transiente* se

$$\sum_{n=1}^{\infty} f_i^{(n)} < 1.$$

Em outras palavras, i é recorrente se, ao partir de i , a cadeia retornará a i com probabilidade 1. Analogamente, i é transiente caso exista probabilidade positiva de que a cadeia não retorne a i se tiver partido de i .

Observe-se ainda que um estado é recorrente se, e somente se, não for transiente e que todo estado absorvente é recorrente, mas não vice-versa. Por exemplo, em uma cadeia ergódica finita todos os estados são recorrentes, mas nenhum é absorvente.

Um outro exemplo importante: no processo de Moran neutro, ver Exemplo 1.6, os estados $1, 2, \dots, N-1$ são todos transientes.

De fato, seja $i \in \{1, 2, \dots, N-1\}$. Sabemos, pela fórmula (1.2.15), que a probabilidade de realizar transições em que o número de indivíduos de tipo A diminua de 1 unidade em uma unidade de tempo é sempre positiva, a não ser quando a cadeia começa em 0 ou N , possibilidades que estamos excluindo aqui. Portanto é positiva a probabilidade de que uma trajetória da cadeia comece no estado i e passe sucessivamente pelos estados $i-1$, $i-2$, \dots , até chegar a 0 em exatas i unidades de tempo. Mais exatamente,

$$P(X_i = 0 | X_0 = i) = \prod_{j=1}^i P(X_j = i-j | X_{j-1} = i-j+1) > 0,$$

pois cada uma das probabilidades no produto à direita é positiva por (1.2.15).

Analogamente, pode-se mostrar que se $i \in \{1, 2, \dots, N - 1\}$, então $P(X_{N-i} = N | X_0 = i) > 0$.

Se a cadeia começa em $i \neq 0, N$, mostramos que com probabilidade positiva chegará ou no estado 0, ou no estado N , ambos absorventes. Portanto existe uma probabilidade *positiva* de que a cadeia que começa em $i \neq 0, N$ não retorne *nunca* a i , pois terá sido absorvida em 0 ou N . Isto prova que os estados $i \in \{1, 2, \dots, N - 1\}$ do processo de Moran neutro são transientes.

Vamos fechar este capítulo resolvendo um mistério pendente desde que definimos o processo de Moran neutro. Isto será feito usando a distribuição geométrica, que revisamos brevemente na seção A.2 do Apêndice. Aqui é também importante que o espaço de estados seja *finito*.

Teorema 1.7. *Seja $(X_n)_{n=0}^\infty$ cadeia de Markov com espaço de estados E finito e seja Ω o espaço amostral de todas as trajetórias da cadeia. Suponha que haja em E um subconjunto não-vazio A de estados absorventes e que os demais estados sejam todos transientes.*

Se $b \in E \setminus A$ é um estado transiente e

$$N_b = \bigcap_{n=1}^{\infty} \{\omega \in \Omega; X_0(\omega) = b, X_n \notin A\}$$

é o conjunto das trajetórias da cadeia que começam em b mas não são absorvidas, então

$$P(N_b) = 0 .$$

Em outras palavras, com probabilidade 1 serão absorvidas todas as trajetórias que começam em um estado transiente.

Demonstração: Se $f_k^{(n)}$ é a probabilidade de que seja no instante n o primeiro retorno a $k \in E$ da cadeia com estado inicial k , então

$$q_k = 1 - \sum_{n=1}^{\infty} f_k^{(n)}$$

é a probabilidade de que a cadeia que começou em k nunca retorne a k . Se b é transiente, então, pela Definição 1.24, $q_b > 0$.

Definindo como um “sucesso” o retorno a b de uma trajetória da cadeia que passou uma primeira vez em b , então a variável aleatória Y_b definida como o número de retornos a b de uma trajetória que passou em b possui distribuição geométrica

$$P(Y_b = n) = p_b^n q_b ,$$

onde $p_b = 1 - q_b$ e $n = 0, 1, 2, \dots$

Calculamos então

$$P(Y_b = \infty) = 1 - \sum_{n=0}^{\infty} p_b^n q_b = 1 - q_b \frac{1}{1 - p_b} = 0,$$

que mostra que as trajetórias que passam por um estado transiente e retornam infinitas vezes a esse estado possuem probabilidade nula.

Observe agora que se ω é trajetória em N_b , então $Y_c(\omega) = \infty$ para algum $c \in E \setminus A$. Caso contrário, como $E \setminus A$ é finito, o número de passagens da cadeia pelos estados de $E \setminus A$ seria finito, contrariando a hipótese de que ω não foi absorvida. Portanto $N_b \subset \bigcup_{c \in E \setminus A} \{Y_c = \infty\}$.

Vice-versa, é fácil ver que $\bigcup_{c \in E \setminus A} \{Y_c = \infty\} \subset N_b$, o que nos faz concluir que $N_b = \bigcup_{c \in E \setminus A} \{Y_c = \infty\}$.

Caso os conjuntos $\{Y_c = \infty\}$ fossem mutuamente disjuntos, teríamos $P(N_b) = \sum_{c \in E \setminus A} P(Y_c = \infty)$, ver (iv) na Definição 1.1. Como isto não é necessariamente verdade, temos que $P(N_b) \leq \sum_{c \in E \setminus A} P(Y_c = \infty) = 0$. Como qualquer probabilidade é sempre não-negativa, então $P(N_b) = 0$. ■

Como o processo de Moran neutro atende às hipóteses do teorema acima, uma vez que há dois estados absorventes e todos os demais $N - 1$ são transientes, então provamos que com probabilidade 1 todas as trajetórias serão absorvidas. No caso de Moran neutro, absorção é fixação, que ocorre portanto com probabilidade 1. Não é que não existam trajetórias que não são absorvidas, mas por terem probabilidade nula, elas são basicamente invisíveis em uma simulação computacional.

Exercícios do capítulo 1

1. Prove a fórmula (1.1.7).

2. A distribuição de Poisson

A distribuição de Poisson é uma distribuição de probabilidades para uma variável aleatória discreta X com espaço de estados $\{0, 1, 2, \dots\}$ e função de probabilidade

$$P(X = k) \equiv p_k = e^{-\lambda} \frac{\lambda^k}{k!},$$

onde $\lambda > 0$ é um parâmetro.

(a) Prove que, naturalmente, $\sum_{k=0}^{\infty} p_k = 1$.

(b) Calcule a média e a variância da distribuição de Poisson.

3. Periodicidade em cadeias de Markov

Considere N pedras em um lago arranjadas de modo a formar um círculo, numeradas de modo que a pedra mais a norte é a número 1 e a pedra à direita da pedra i tem número $i + 1$, $i = 1, 2, \dots, N - 1$. A cada unidade de tempo um sapo joga uma moeda e salta para a pedra à sua esquerda se sair cara, e para a pedra à sua direita se sair coroa. Suponha que no instante 0 o sapo esteja na pedra 1 e que a moeda não é necessariamente honesta, saindo em cara com probabilidade p e coroa com probabilidade $1 - p$.

- (a) Mostre que se N é par, então o sapo estará em uma pedra par nos instantes ímpares e numa pedra ímpar nos instantes pares. Mostre que se N é ímpar, é impossível prever a paridade da pedra em que estará o sapo.
- (b) Seja X_n a posição do sapo no instante n . Mostre que $(X_n)_{n=0}^{\infty}$ é uma cadeia de Markov e escreva sua matriz de transição T . Mostre que T é irredutível.
- (c) Mostre que T , além de estocástica por colunas, também é estocástica por linhas. Conclua que $\mathbf{1}$ é autovetor de T com autovalor 1.
- (d) Prove que se N é par, então o vetor cujas componentes alternam entre -1 e 1 é autovetor de T com autovalor -1 .
- (e) Use a informação do item (a) para provar que se N é par existem vetores de probabilidade w tais que se w é a distribuição de probabilidade para X_0 , então $T^n w$ não possui limite quando $n \rightarrow \infty$.
- (f) É possível provar que os resultados neste item valem para qualquer N ímpar, mas propomos aqui estudar somente o caso $N = 3$. Se $N = 3$, mostre que o único autovalor de T com módulo 1 é 1. Conclua que para qualquer distribuição de probabilidade w para X_0 existirá sempre $\lim_{n \rightarrow \infty} T^n w$. Quanto vale tal limite?
4. Considere a cadeia de Markov com estados $\{1, 2, 3, 4\}$ e a matriz de transição

$$M = \begin{pmatrix} \frac{p}{2} & 0 & \frac{p}{2} & 0 \\ \frac{p}{2} & 0 & \frac{p}{2} & 0 \\ \frac{1-p}{2} & \frac{1}{2} & \frac{1-p}{2} & \frac{1}{2} \\ \frac{1-p}{2} & \frac{1}{2} & \frac{1-p}{2} & \frac{1}{2} \end{pmatrix},$$

onde $p \in [0, 1]$.

- (a) Mostre que os autovalores de M são 0 (com multiplicidade 3) e 1, independentemente de p .
- (b) Calcule a distribuição estacionária da cadeia quando $p = 1/4$.
- (c) Calcule a distribuição estacionária da cadeia quando $p = 0$.
- (d) Mostre que cadeia é irredutível se $p > 0$.
- (d) Se $p = 0$ a cadeia é redutível. Ainda assim pode-se afirmar que a distribuição estacionária é a distribuição limite para qualquer condição inicial?

5. Número de caminhos em um digrafo

Seja D um digrafo com vértices $\{1, 2, \dots, n\}$ e A sua matriz de adjacências. Use indução para provar que o número de caminhos de comprimento k que liga o vértice j ao vértice i é o elemento ij da matriz A^k .

6. Seja P matriz $n \times n$ não-negativa. Então, se $p_{ij}^{(n)}$ denotam os elementos de P^n , prove que

$$p_{ij}^{(n)} > 0 \Leftrightarrow \text{existe caminho de comprimento } n \text{ ligando } j \text{ a } i \text{ no digrafo de adjacências de } P.$$

7. Use o exercício anterior para provar a Proposição 1.2.

8. O processo de Moran neutro com mutação unidirecional

Considere a seguinte modificação do processo de Moran neutro, descrito no Exemplo 1.6: o descendente de um indivíduo de tipo A possui probabilidade $\mu > 0$ de ser de tipo B (por uma mutação genética) e probabilidade $1 - \mu$ de ser de tipo A . Os descendentes de indivíduos de tipo B , ao contrário, são sempre de tipo B . As variáveis aleatórias que nos interessa descrever (estado da população) são os números de indivíduos do tipo A em cada instante de tempo.

- (a) Encontre uma expressão para os elementos da matriz de transição da cadeia de Markov que descreve esse processo de Moran modificado.
- (b) Classifique cada um dos estados da cadeia como transientes ou recorrentes.
- (c) O que se deve esperar do estado final dessa população?

9. O processo de Moran neutro com mutação bidirecional

Considere esta outra modificação do processo de Moran neutro: o descendente de um indivíduo de tipo A possui probabilidade $\mu_1 > 0$ de ser de tipo B e probabilidade $1 - \mu_1$ de ser de tipo A e o descendente de um indivíduo de tipo B possui probabilidade $\mu_2 > 0$ de ser de tipo A e probabilidade $1 - \mu_2$ de ser de tipo B . As variáveis aleatórias que nos interessa descrever são os números de indivíduos do tipo A em cada instante de tempo.

- (a) Encontre uma expressão para os elementos da matriz de transição da cadeia de Markov que descreve esse processo.
- (b) Classifique cada um dos estados da cadeia como transientes ou recorrentes.
- (c) O que se deve esperar do estado final dessa população?

Capítulo 2

O processo de Moran generalizado

Neste capítulo vamos generalizar o processo de Moran neutro introduzido no Exemplo 1.6 a situações onde não se tenha necessariamente a neutralidade, ou seja, onde os indivíduos de tipo A ou B não são necessariamente equivalentes do ponto de vista da seleção natural. Na verdade vamos generalizar ainda mais, colocando os processos de Moran na classe mais ampla conhecida como *processos de nascimento e morte*, que inclui também o famoso *problema da ruína do jogador*.

Tendo aprendido ao final do capítulo anterior que no processo de Moran neutro a fixação ocorre com probabilidade 1, vamos nos preocupar agora com duas questões quantitativas: dada a frequência inicial de indivíduos de tipo A na população

1. Qual a probabilidade de fixação em cada um dos dois estados absorventes?
2. Quanto tempo em média deve-se esperar até que ocorra a fixação?

Veremos que o formalismo necessário para responder a essas perguntas pode ser formulado de maneira geral e depois simplificado no caso de processos de nascimento e morte.

Lembramos que o processo de Moran é um *modelo* para a evolução biológica de uma população do ponto de vista dos genes desta população. Como todo modelo, o processo de Moran é uma tentativa de descrição da realidade a partir de simplificações introduzidas. Em particular, as versões que estudaremos são as mais simples possíveis, embora, como veremos, já

sejam suficientemente ricas para exibir fenômenos interessantes. Em particular, vamos supor, em todas as versões que estudarmos para o processo de Moran, que:

1. O tempo passa em unidades discretas.
2. A população total é constante igual a N indivíduos.
3. A reprodução é assexuada, pois cada indivíduo possui somente um ancestral.
4. Não existe nenhuma estrutura na população, seja esta espacial, social, ou de idades.

A remoção de quaisquer dessas suposições leva a problemas bastante interessantes.

Uma outra característica do processo de Moran, como veremos nas próximas seções, é a possibilidade de realizar exatamente os cálculos das probabilidades de absorção em cada um dos estados absorventes e dos tempos médios para a absorção.

2.1 Processos de nascimento e morte

Se $\alpha_i, \beta_i, i = 1, 2, \dots, N-1$ são parâmetros em $[0, 1]$ e tais que $\alpha_i + \beta_i \leq 1$, um processo de nascimento e morte é definido como uma cadeia de Markov com estados $E = \{0, 1, 2, \dots, N\}$ e matriz de transição

$$T = \begin{pmatrix} 1 & \beta_1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 - \alpha_1 - \beta_1 & \beta_2 & \cdots & 0 & 0 & 0 \\ 0 & \alpha_1 & 1 - \alpha_2 - \beta_2 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \alpha_2 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 - \alpha_{N-2} - \beta_{N-2} & \beta_{N-1} & 0 \\ 0 & 0 & 0 & \cdots & \alpha_{N-2} & 1 - \alpha_{N-1} - \beta_{N-1} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \alpha_{N-1} & 1 \end{pmatrix}.$$

Aqui é preciso atentar para o fato de que o primeiro estado em E é o estado 0 e portanto a primeira coluna de T representa as probabilidades de transição do estado 0 para os estados de E . Conseqüentemente, a segunda coluna de T são as probabilidade de transição para o estado 1 e assim por diante. Em particular, T é $(N + 1) \times (N + 1)$ e é tridiagonal, ou seja, os elementos não-nulos de T ocorrem somente na diagonal principal e nas diagonais imediatamente acima e abaixo desta.

Observe que a matriz de transição do processo de Moran neutro, cujos elementos são dados por (1.2.15) possui exatamente a forma da matriz acima, com

$$\alpha_i = \beta_i = \frac{i}{N} \frac{N-i}{N}, \quad (2.1.1)$$

$i = 0, 1, 2, \dots, N$.

Portanto o processo de Moran neutro é um exemplo de processo de nascimento e morte. Mais em geral, podemos pensar um processo de nascimento e morte como um processo que descreve uma população de tamanho fixo igual a N indivíduos, em que os indivíduos são de dois tipos – digamos tipos A e B – e a variável aleatória X_n representa o número de indivíduos de tipo A no instante $n = 0, 1, 2, \dots$. Segundo a forma tridiagonal da matriz de transição, a cada unidade de tempo o número de indivíduos de cada tipo ou permanece constante, ou cresce de uma unidade, ou decresce de uma unidade. O parâmetro α_i é interpretado como a probabilidade de aumento dos indivíduos de tipo A (e portanto diminuição dos de tipo B) quando o número destes é igual a i e, similarmente, β_i é a probabilidade de diminuição de indivíduos de tipo A quando o número destes é igual a i .

Um outro exemplo interessante é o chamado *problema da ruína do jogador*, um dos clássicos problemas da Teoria de Probabilidades. Nesse caso, tomamos $\alpha_i = p$, $\beta_i = q$, com $p, q \in [0, 1]$, $p + q = 1$. A interpretação do problema da ruína do jogador é que a cada unidade de tempo o jogador 1 e o jogador 2 jogam uma determinada partida (moeda, dado, roleta, etc), em que com probabilidade p o jogador 1 ganha e com probabilidade q o jogador 1 perde. Supomos que os jogadores apostam uma unidade de dinheiro a cada partida, que o capital somado dos dois é igual a N , que o jogo só termina quando um dos dois jogadores fica sem dinheiro e que X_n representa o capital do jogador 1 no instante n .

Nos processos de nascimento e morte os estados 0 e N são claramente absorventes, basta olhar para a primeira e a última colunas de T , e se α_i e β_i são positivos, todos os demais estados são transientes. O raciocínio para provar isto é o mesmo que já usamos para a mesma afirmativa no caso do processo de Moran neutro. Valem portanto as hipóteses do Teorema 1.7 e portanto podemos dizer em geral que com probabilidade 1 o processo será absorvido ou no estado 0, ou no estado N . No caso de estarmos interpretando X_n como número de indivíduos de tipo A vemos que com probabilidade 1 haverá extinção ou dos indivíduos de tipo A, ou dos de tipo B. No caso da ruína do jogador vemos que o jogo irá terminar com probabilidade 1 em tempo finito.

Perguntas naturais a que iremos responder: dado o número inicial de indivíduos de tipo A, qual a probabilidade de extinção do tipo B e quanto

tempo leva essa extinção? Analogamente, dado o capital inicial do jogador 1, qual a probabilidade de que este arruíne o jogador 2 e em quanto tempo ele o fará?

Um caso interessante é quando o jogador de tipo 2 é um cassino e – naturalmente – o capital inicial k do jogador 1 é muito menor que o capital $N - k$ do cassino. Será que mesmo em condições de jogo honesto ($p = q = \frac{1}{2}$) há uma boa chance de “quebrar a banca”? Ao respondermos a esta pergunta o leitor talvez considerará um bom investimento abrir um cassino. Devemos lembrar-lhe que o jogo é ilegal no Brasil e, ademais, é preciso também ter um bom capital inicial. Vamos porém nos divertir com o cálculo.

Antes de passar à resposta a essas perguntas podemos ainda remover a hipótese de neutralidade do processo de Moran, escondida em (2.1.1) no fato de que são iguais as probabilidades de aumento e diminuição de indivíduos de tipo A. Podemos supor por exemplo que as mortalidades de indivíduos de tipo A e B são iguais, mas não suas natalidades. Somos levados então a uma definição da maior importância biológica.

Definição 2.1. *A aptidão absoluta (ou, comumente, o termo em inglês fitness absoluta) de um indivíduo é o número de descendentes que esse indivíduo deixa na geração seguinte.*

Muitas vezes, em vez de aptidão absoluta usa-se a aptidão relativa, em que o número de descendentes de um indivíduo é especificado não de forma absoluta, mas em unidades do número de descendentes de algum indivíduo padrão.

Vamos supor que a aptidão média dos indivíduos da população dependa somente do tipo do indivíduo. Escolhendo como padrão os indivíduos de tipo B, aos quais atribuiremos aptidão relativa igual a 1, seja r a aptidão relativa dos indivíduos de tipo A. Caso $r > 1$, os indivíduos de tipo A têm em média mais descendentes que os de tipo B. Caso $0 < r < 1$, os indivíduos de tipo B terão em média mais descendentes que os de tipo A.

Para acomodar a diferença de aptidão dos dois tipos, mas mantendo iguais as mortalidades, vamos modificar as regras para a simulação do processo de Moran descritas no Exemplo 1.6. A cada unidade de tempo vamos continuar escolhendo de maneira *uniforme* um indivíduo para morrer. Vamos também escolher o tipo de um único indivíduo que irá nascer, substituindo o que morreu. Em vez de escolher como antes, vamos supor que produzamos r descendentes virtuais para cada indivíduo A e 1 descendente virtual para cada indivíduo B. O novo indivíduo que nascerá na população será escolhido de maneira uniforme entre os $ri + N - i$ indivíduos virtuais, onde i é o número de indivíduos de tipo A. Portanto, a probabilidade de

que seja sorteado para nascer um indivíduo de tipo A é

$$\frac{ri}{ri + N - i}, \quad (2.1.2)$$

e a probabilidade de nascer um de tipo B é

$$\frac{N - i}{ri + N - i}.$$

Com esta regra, o processo de Moran com aptidão r para indivíduos de tipo A é definido pelos parâmetros

$$\alpha_i = \frac{ri}{ri + N - i} \frac{N - i}{N} \quad \text{e} \quad \beta_i = \frac{N - i}{ri + N - i} \frac{i}{N}, \quad (2.1.3)$$

conforme o leitor pode verificar facilmente.

2.2 Probabilidade de fixação

Suponha que temos uma cadeia de Markov com matriz de transição T e estados $0, 1, \dots, N$, sendo os estados 0 e N absorventes e todos os demais, transientes. Já sabemos pelo Teorema 1.7 que a cadeia será absorvida com probabilidade 1 ou no estado 0 , ou no estado N . Podemos então definir um vetor $\pi = (\pi_0, \pi_1, \dots, \pi_N) \in \mathbb{R}^{N+1}$, em que π_i é a probabilidade de absorção da cadeia no estado N , *dado que o estado inicial é i* . Obviamente, a probabilidade de fixação da cadeia no estado 0 , dado o estado inicial i é $1 - \pi_i$. Nosso objetivo é encontrar uma equação geral para determinar o vetor π das probabilidades de absorção. Em seguida veremos que no caso dos modelos de nascimento e morte é possível resolver essa equação explicitamente.

Notamos primeiramente que

$$\pi_0 = 0 \quad \text{e} \quad \pi_N = 1. \quad (2.2.4)$$

Em seguida, pensemos às trajetórias da cadeia que contribuem para a probabilidade π_i . Elas se encontram com certeza no estado i a tempo 0 , mas não sabemos com certeza onde estarão no tempo 1 : podem estar com probabilidade t_{ji} no estado j e, caso estejam em j , a probabilidade de serem no futuro absorvidas em N é π_j . Podemos encontrar π_i somando sobre todas as possibilidades j para o estado da cadeia no tempo 1 , cada uma pesada pela probabilidade t_{ji} de realizar a transição do estado i no tempo 0 para o estado j no tempo 1 . Ou seja, temos

$$\pi_i = \sum_{j=0}^N t_{ji} \pi_j, \quad (2.2.5)$$

que vale para $i = 1, 2, \dots, N - 1$ e deve ser acompanhada nos casos das condições de contorno (2.2.4).

É interessante notar que as equações acima para $i = 0, 1, \dots, N$ podem ser reescritas conjuntamente em notação matricial como

$$\pi = T^t \pi, \quad (2.2.6)$$

mostrando que π pode ser calculado como autovetor de T^t com autovalor 1, sendo que (2.2.4) entra como uma condição que seleciona de todo o autoespaço o vetor adequado. Observe que já sabíamos do Teorema 1.1 que T^t possui 1 como autovalor, sendo interessante encontrarmos aqui uma interpretação para um dos autovetores correspondentes. Observe ainda que, como temos dois estados absorventes, que são naturalmente autovetores de T com autovalor 1, então o autoespaço de T^t com autovalor 1 tem que ser pelo menos bidimensional. Como nenhum múltiplo do autovetor $\mathbb{1}$ de T^t encontrado no Teorema 1.1 satisfaz (2.2.4), é natural que tenha que existir pelo menos uma outra dimensão no autoespaço de T^t para acomodar a solução para (2.2.5) e (2.2.4).

Até agora trabalhávamos na hipótese geral de que T era matriz de transição de uma cadeia de Markov com estados $0, 1, \dots, N$, sendo os estados 0 e N absorventes e todos os demais, transientes. Somente agora particularizaremos para o caso dos processos de nascimento e morte. Nesse caso, somente três dos termos da soma em (2.2.5) são possivelmente não-nulos e tal simplificação nos leva a uma fórmula explícita para os π_i . No caso de nascimento e morte, a equação (2.2.5) torna-se

$$\pi_i = \beta_i \pi_{i-1} + (1 - \alpha_i - \beta_i) \pi_i + \alpha_i \pi_{i+1}, \quad (2.2.7)$$

que reescrevemos como

$$\alpha_i (\pi_{i+1} - \pi_i) - \beta_i (\pi_i - \pi_{i-1}) = 0.$$

Definindo $x_i = \pi_i - \pi_{i-1}$ encontramos que

$$x_{i+1} = \gamma_i x_i,$$

$i = 1, 2, \dots, N - 1$, onde definimos ainda $\gamma_i = \frac{\beta_i}{\alpha_i}$.

Em termos das novas variáveis x_i a solução agora é fácil:

$$\begin{aligned} x_1 &= \pi_1 \\ x_2 &= \gamma_1 \pi_1 \\ x_3 &= \gamma_2 \gamma_1 \pi_1 \\ &\vdots \end{aligned}$$

Em geral,

$$x_n = \left(\prod_{k=1}^{n-1} \gamma_k \right) \pi_1, \quad (2.2.8)$$

$n = 2, 3, \dots, N$

Por um lado, temos, pela definição dos x_i ,

$$\sum_{i=1}^N x_i = \sum_{i=1}^N (\pi_i - \pi_{i-1}) = \pi_N - \pi_0 = 1,$$

onde usamos (2.2.4). Por outro, temos

$$\sum_{i=1}^N x_i = \left(1 + \sum_{j=1}^{N-1} \prod_{k=1}^j \gamma_k \right) \pi_1.$$

Igualando as duas expressões e resolvendo para x_1 temos

$$\pi_1 = x_1 = \frac{1}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^j \gamma_k}. \quad (2.2.9)$$

Reescrevendo os π_n em termos dos x_n e usando sucessivamente (2.2.8) encontraremos,

$$\pi_2 = \pi_1 + x_2 = (1 + \gamma_1)\pi_1,$$

$$\pi_3 = \pi_2 + x_3 = (1 + \gamma_1 + \gamma_1\gamma_2)\pi_1,$$

e, em geral,

$$\pi_i = \left(1 + \sum_{j=1}^{i-1} \prod_{k=1}^j \gamma_k \right) \pi_1. \quad (2.2.10)$$

Finalmente, substituindo (2.2.9) em (2.2.10) obtemos

$$\pi_i = \frac{1 + \sum_{j=1}^{i-1} \prod_{k=1}^j \gamma_k}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^j \gamma_k}, \quad (2.2.11)$$

que vale para $i = 2, 3, \dots, N - 1$ e, junto com (2.2.4) e (2.2.9), determinam completamente o vetor π .

Apesar de explícitas, as fórmulas que encontramos não são particularmente transparentes. Veremos que ao substituirmos nelas os valores de γ_i para as situações de interesse e realizarmos simplificações teremos resultados facilmente interpretáveis.

Tanto no caso do processo de Moran *neutro* quanto no da ruína do jogador *neutra*, ou seja com $p = q = 1/2$, temos $\gamma_i = 1$ e portanto

$$\pi_i = \frac{i}{N},$$

$i = 0, 1, \dots, N$. No Moran neutro, a probabilidade de que os indivíduos de tipo B sejam extintos é igual à fração inicial de indivíduos de tipo A na população. Na ruína do jogador neutra a probabilidade de que o jogador 2 saia arruinado é a fração inicial do capital total possuída pelo jogador 1. Como, em geral, o capital inicial i do jogador 1 é muito menor que o capital do cassino $N - i$, é claro que é muito improvável arruinar um cassino, mesmo no caso de jogo neutro.

Nenhuma das conclusões é muito surpreendente até aqui, mas ambas são razoáveis, indicando a correção do cálculo. No caso da ruína do jogador não-neutra, se p é a probabilidade de o jogador 1 ganhar cada partida e $q = 1 - p$ é a probabilidade do cassino, então $\gamma_i = \frac{q}{p}$ e temos, usando a bem conhecida fórmula

$$1 + x + x^2 + \dots + x^{n-1} = \frac{x^n - 1}{x - 1}$$

para a soma de um número finito de termos de uma progressão geométrica, que

$$\pi_i = \frac{\left(\frac{q}{p}\right)^i - 1}{\left(\frac{q}{p}\right)^N - 1}.$$

Se o cassino tiver uma ligeira vantagem em cada partida, que é o que acontece por exemplo em um jogo de roleta, e i for muito menor que N , é virtualmente impossível arruinar o cassino. Podemos por exemplo desprezar 1 no numerador e no denominador, desde que $q > p$ e i seja grande o suficiente, e obtemos $\pi_i \approx (q/p)^{i-N}$. Na Figura 2.1 mostramos a probabilidade de arruinar um cassino em função do capital inicial i tanto no caso neutro, quanto no caso $p = 0,49, q = 0,51$ em que o cassino tem uma ligeira vantagem.

No caso do processo de Moran não-neutro, uma situação interessante é a de calcular a probabilidade de sobrevivência em uma população de tamanho N de uma mutação vantajosa. As mutações são geralmente muito raras e aparecem inicialmente em um único indivíduo. Digamos que um único indivíduo de tipo A aparece em uma população com mais $N - 1$ indivíduos de tipo B. Supomos que a aptidão de A relativa a B seja r . Neste caso, de acordo com a discussão do processo de Moran não-neutro que fizemos mais

Figura 2.1: Probabilidade de arruinar um cassino como função do capital inicial do jogador. O gráfico acima é o do caso de jogo neutro, enquanto o de baixo é com vantagem ligeira para o cassino, em que tomamos $p = 0,49, q = 0,51$. Em ambos os casos, o capital total do jogo é $N = 1000$.

atrás, teremos $\gamma_i = 1/r$. O cálculo é formalmente idêntico ao da ruína do jogador, só que agora nos interessa fixar o número inicial de indivíduos de tipo A em 1 e olhar para π_1 como função da população total N :

$$\pi_1 = \frac{1 - \frac{1}{r}}{1 - \left(\frac{1}{r}\right)^N}.$$

A probabilidade de fixação de um mutante vantajoso $r > 1$ tende a $1 - \frac{1}{r}$ quando $N \rightarrow \infty$. Na Figura 2.2 mostramos como função do tamanho da população a probabilidade de fixação π_1 de um mutante que inicialmente apareça uma única vez e tenha vantagem de aptidão de 20% com relação resto da população.

O caso $0 < r < 1$, em que a mutação do indivíduo de tipo A é desvantajosa, corresponde ao caso da ruína do jogador com $q > p$: se N for grande é virtualmente impossível que uma mutação mesmo pouco desvantajosa se estabeleça em uma população. Caso a população não seja assim tão grande, mesmo uma mutação desfavorável pode fixar-se. Teremos mais a comentar a respeito de populações pequenas mais adiante.

No caso $r > 1$, a mutação é vantajosa e vemos que, mesmo com a vantagem de maior número médio de descendentes, não é garantido que haverá fixação de A. Por outro lado, caso um mutante com uma mutação vanta-

Figura 2.2: Probabilidade de fixação de um mutante vantajoso, $r = 1, 2$, como função do tamanho N da população no processo de Moran. Observe que para populações grandes a probabilidade π_1 fica próxima de 17%.

josa consiga prosperar e atingir uma determinada proporção da população, digamos 10%, então será praticamente impossível que ele não se fixe. Na Figura 2.3 mostramos a probabilidade de fixação como função do tamanho da população para uma mutação vantajosa com $r = 1, 2$, mesmo valor usado na Figura 2.2, que já tenha alcançado 10% do valor da população. Como se vê, uma vez atingida tal proporção, em uma população grande o suficiente a sobrevivência de uma mutação vantajosa é praticamente certa.

2.3 Tempo médio de fixação

Nesta seção vamos repetir o roteiro da última: primeiramente deduzimos uma fórmula geral para o cálculo do tempo de absorção e depois a aplicamos ao caso especial e mais simples de um processo de nascimento e morte. Em seguida exploraremos as consequências dos cálculos.

Suponhamos novamente uma cadeia de Markov com matriz de transição T e estados $0, 1, \dots, N$, sendo os estados 0 e N absorventes e todos os demais, transientes. Definimos agora um vetor $\tau = (\tau_0, \tau_1, \dots, \tau_N) \in \mathbb{R}^{N+1}$, em que τ_i é o tempo *médio* de absorção da cadeia, *dado que o estado inicial é i* , onde não se especifica em qual dos dois estados será a absorção.

Assim como no caso das probabilidades de fixação, começamos com as

Figura 2.3: Probabilidade de fixação de uma mutação vantajosa com $r = 1, 2$ como função do tamanho N da população no processo de Moran, considerando que inicialmente 10% da população seja constituída por esses mutantes.

condições de contorno:

$$\tau_0 = 0 \quad \text{e} \quad \tau_N = 0. \quad (2.3.12)$$

Suponhamos agora uma trajetória da cadeia começando no estado i e com tempo médio de absorção τ_i . No instante 1 ela estará no estado j com probabilidade t_{ji} e a partir de então o tempo médio de absorção será τ_j , pois a cadeia já terá perdido a memória do estado inicial (propriedade de Markov). A equação abaixo, análoga a (2.2.5), cristaliza este raciocínio:

$$\tau_i = 1 + \sum_{j=0}^N t_{ji} \tau_j, \quad (2.3.13)$$

$i = 1, 2, \dots, N - 1$. O termo 1 somado no início do segundo membro se deve ao fato de que do lado direito da equação estamos contando tempos de absorção a partir do tempo 1 e, portanto, para comparar com o tempo contado no lado esquerdo devemos somar 1.

Uma maneira prática para se calcular τ em algumas situações é simplesmente resolver (2.3.13) como um sistema de equações lineares cuja forma matricial é

$$(T^t - I)\tau = -(0, 1, \dots, 1, 0)^t,$$

onde I denota, como usual, a matriz identidade. É interessante notar que sabemos já que a matriz $T^t - I$ dos coeficientes do sistema é singular. No

entanto o sistema possui solução e, portanto, infinitas soluções. As condições de contorno (2.3.12) devem selecionar a solução única para determinar os tempos médios de absorção.

No caso de processos de nascimento e morte a equação (2.3.13) torna-se

$$\tau_i = 1 + \beta_i \tau_{i-1} + (1 - \alpha_i - \beta_i) \tau_i + \alpha_i \tau_{i+1}, \quad (2.3.14)$$

ou melhor,

$$\alpha_i(\tau_{i+1} - \tau_i) - \beta_i(\tau_i - \tau_{i-1}) = -1.$$

Definindo $z_i = \tau_i - \tau_{i-1}$ encontramos que

$$z_{i+1} = \gamma_i z_i - \frac{1}{\alpha_i},$$

$i = 1, 2, \dots, N-1$, onde $\gamma_i = \frac{\beta_i}{\alpha_i}$ já tinha sido introduzido na seção anterior.

Usando recursivamente a relação entre z_{i+1} e z_i , uma fácil indução mostra que se $n = 2, 3, \dots, N$

$$z_n = \left(\prod_{j=1}^{n-1} \gamma_j \right) z_1 - \sum_{j=1}^{n-1} \frac{1}{\beta_j} \prod_{k=j}^{n-1} \gamma_k. \quad (2.3.15)$$

Como $\sum_{n=1}^N z_n = \tau_N - \tau_0 = 0$, então

$$\tau_1 = z_1 = \frac{\sum_{n=2}^N \sum_{j=1}^{n-1} \frac{1}{\beta_j} \prod_{k=j}^{n-1} \gamma_k}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^j \gamma_k}. \quad (2.3.16)$$

Fórmulas para z_2, \dots, z_N e, conseqüentemente, $\tau_2, \dots, \tau_{N-1}$ podem ser obtidas substituindo (2.3.16) em (2.3.15), mas o resultado é bem mais complicado que o cálculo correspondente das probabilidades de absorção. O motivo dessa diferença é que tanto a fórmula (2.3.16), quanto a relação (2.3.15) dependem também dos β_j , ao contrário de seus análogos da seção anterior que só dependem dos γ_j . Vamos estudar aqui somente o caso do processo de Moran neutro e o da ruína do jogador não-neutra. O caso da ruína do jogador neutro será deixado como exercício.

No caso do processo de Moran neutro, vale (2.1.1), logo $\gamma_i = 1$ e portanto (2.3.16) torna-se, após fáceis simplificações,

$$\tau_1 = N \sum_{n=2}^N \sum_{j=1}^{N-1} \frac{1}{j(N-j)}.$$

Fazendo a decomposição em frações parciais $\frac{1}{j(N-j)} = \frac{1}{N} \left(\frac{1}{j} + \frac{1}{N-j} \right)$ e colecionando termos idênticos na soma dupla, chegamos a

$$\tau_1 = N h_{N-1},$$

onde

$$h_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$$

é a soma parcial da *série harmônica*.

A divergência da série harmônica [?] é fato bem conhecido desde os primeiros cursos de Cálculo e este fato aqui tem uma consequência muito importante: τ_1 (e também os demais tempos médios de absorção) divergem no limite de população infinita. Este fato é tão importante para a Genética de Populações que será melhor explorado na próxima seção. Deixemo-lo portanto momentaneamente de lado e continuemos nossos cálculos.

Continuando os cálculos e com um pouco de trabalho chega-se à seguinte fórmula para os tempos de absorção para o processo de Moran neutro:

$$\begin{aligned} \tau_n &= N \left[(N-n) \sum_{j=1}^n \frac{1}{N-j} + n \sum_{j=n+1}^{N-1} \frac{1}{j} \right] \\ &= N \left[(N-n)(h_{N-1} - h_{N-n-1}) + n \sum_{j=n+1}^{N-1} (h_{N-1} - h_n) \right] \end{aligned} \quad (2.3.17)$$

onde nota-se novamente o aparecimento de somas parciais da série harmônica.

Para entendermos melhor o resultado acima podemos recorrer à chamada fórmula de Euler-MacLaurin [?], que fornece uma *aproximação* para h_n *assintótica* no limite $n \rightarrow \infty$:

$$h_n \approx \log n + \gamma + \frac{1}{2n}, \quad (2.3.18)$$

onde o logaritmo é naturalmente o logaritmo natural e γ denota aqui a chamada *constante de Euler-Mascheroni*, $\gamma \approx 0,577$.

Substituindo a aproximação (2.3.18) na fórmula exata para τ_n e desprezando termos do tipo $\frac{1}{n}$, muito menores que os logaritmos, obtemos

$$\begin{aligned} \tau_n &\approx N \left[(N-n) \log \frac{N-1}{N-1-n} + n \log \frac{N-1}{n-1} \right] \\ &\approx -N^2 [x \log x + (1-x) \log(1-x)], \end{aligned} \quad (2.3.19)$$

em que introduzimos na última linha a fração inicial $x = \frac{n}{N}$ de indivíduos de tipo A na população total.

A Figura 2.4 mostra um gráfico do tempo médio de absorção na aproximação (2.3.19) como função de x para uma população de $N = 50$ indivíduos. Para comparação, mostramos também o gráfico do tempo exato (2.3.17). Como a aproximação é assintótica no limite $N \rightarrow \infty$, para valores

Figura 2.4: Tempo médio de absorção no processo de Moran neutro com $N = 50$ como função da fração inicial x de indivíduos de tipo A. A curva cheia é o gráfico da solução exata (2.3.17) e a tracejada é o gráfico da aproximação (2.3.19).

maiores de N fica ainda mais difícil notar a diferença entre os valores exatos e os aproximados.

A aproximação (2.3.19) deixa claro, conforme já tínhamos mencionado antes, que os tempos médios de absorção no processo de Moran neutro tendem a infinito no limite de população N infinita. Aqui fica patente ainda que os tempos são proporcionais ao *quadrado* da população. Outro ponto relevante é que é fácil provar que o valor máximo da função $-[x \log x + (1-x) \log(1-x)]$ em $(0, 1)$ ocorre em $x = 1/2$ e vale cerca de $0,69$, de forma que os maiores tempos de absorção são cerca de $0,69N^2$.

No caso não-neutro da ruína do jogador, se p é a probabilidade de vitória do jogador 1 em cada partida e $q = 1 - p$ é a probabilidade de vitória do jogador 2, temos $\gamma_i = \frac{q}{p}$, $\beta_i = q$. Os cálculos são trabalhosos, mas obtemos ao final que o tempo médio para a ruína de um dos dois jogadores é

$$\tau(x) = \frac{N}{p(\gamma - 1)} \left(x - \frac{\gamma^{Nx} - 1}{\gamma^N - 1} \right), \quad (2.3.20)$$

onde $\gamma = q/p$, N é o capital total e $x \in [0, 1]$ é a fração do capital total possuída inicialmente pelo jogador 1. Mostramos na Figura 2.5 o gráfico dessa função usando os mesmos parâmetros da parte não-neutra na Figura 2.1: $p = 0.49$, $q = 0.51$, $N = 1000$. Observe que enquanto x é pequeno, segundo a Figura 2.1 é quase impossível que o jogador 1 quebre o jogador 2, o cassino. Na Figura 2.5 vemos que na mesma região de valores de

x tem-se que o termo $\frac{\gamma^{Nx}-1}{\gamma^N-1}$ é desprezível com relação a x e o gráfico é aproximadamente uma reta. Portanto, quando o capital inicial do jogador 1 é pequeno com relação ao do jogador 2, o tempo médio para a falência do jogador 1 é proporcional a N .

Figura 2.5: Tempo médio de duração do processo da ruína do jogador como função da fração do capital total inicialmente possuída pelo jogador 1. Os parâmetros são $p = 0.49$, $q = 0.51$, $N = 1000$.

2.4 Populações pequenas e deriva genética

Conforme mencionamos na exposição do processo de Moran neutro, uma primeira e ingênua expectativa é que se os indivíduos de tipo A e B possuem a mesma aptidão, ou seja, a seleção natural não age nem a favor, nem contra nenhum dos dois tipos, então em uma população onde ambos os tipos estejam presentes, a frequência de cada tipo deveria permanecer aproximadamente constante ao longo do tempo. Através das simulações computacionais do modelo, Figura 1.2, vimos que a constância das frequências não acontecia. Pelo contrário, as frequências flutuavam bastante a ponto de um dos dois tipos acabar extinto. Vimos em seguida que a fixação de um dos dois tipos na população, fixação essa observada nas simulações computacionais, é que é o comportamento típico do modelo, ocorrendo com probabilidade 1, ver Teorema 1.7. Fomos capazes de calcular a probabilidade de fixação de cada um dos tipos como função de sua frequência inicial na população e até o tempo médio decorrido até a fixação.

Um resultado importante foi a *divergência* dos tempos médios de fixação no processo de Moran neutro quando o tamanho da população N tende ao infinito, ver equação (2.3.19). Esse resultado deixa claro que a absorção com probabilidade 1 de uma cadeia de Markov é um *fenômeno de população finita*, algo que também enfatizamos na demonstração do Teorema 1.7.

Obviamente todas as populações existentes são finitas e o fato de não observarmos fixação de caracteres neutros em todas elas poderia ser apontado como uma falha do processo de Moran em modelar a evolução genética de uma população. É verdade que o processo de Moran é sim uma simplificação extremada do que ocorre na natureza, mas podemos creditar a discrepância entre o modelo e a realidade – nem sempre há homogeneidade de tipos em uma população – não à simplicidade extremada do modelo, mas ao fato de que as populações observadas talvez sejam grandes demais para que a fixação tenha ocorrido, lembremo-nos sempre de (2.3.19). E se desejarmos estudar populações grandes por longos períodos de tempo, temos também que estar cientes de que o processo de Moran já não é suficiente para descrevê-las: teríamos que levar em conta várias “forças” genéticas ainda não consideradas: crescimento populacional, emigração e imigração e, principalmente, a ocorrência de mutações genéticas.

O fato de que as frequências gênicas em uma população flutuam mesmo em condições de neutralidade, ou seja, ausência de uma “força” de seleção natural, é conhecido em Genética como *deriva genética*. Se quisermos observar na natureza exemplos de fixação por deriva genética, segundo os argumentos acima temos que procurá-los entre populações pequenas o suficiente para que a fixação tenha tido tempo de ocorrer. Exemplos citados pelos geneticistas são os de populações em ilhas, que são frequentemente pequenas e permanecem isoladas por longos períodos de tempo. Existem documentos fósseis de elefantes anões que teriam existido em ilhas do sudeste da Ásia e do Mediterrâneo e até mesmo de uma população hominídea anã que teria vivido na ilha de Flores, na Indonésia. Tais exemplos de populações em ilhas com características bastante diferentes da população como um todo podem ser consequências de deriva genética.

Exercícios do capítulo 2

1. Crescimento dos π_i para processos de nascimento e morte.
 - (a) Reescreva a equação (2.2.7) e conclua que se $i = 1, 2, \dots, N - 1$, π_i é uma determinada *média ponderada* de π_{i-1} e π_{i+1} .
 - (b) Usando a informação do item anterior e as condições de contorno (2.2.4) prove que a sequência das componentes π_i do vetor π é

crescente no índice i .

2. Princípios do máximo e do mínimo.

- (a) Observe que, como no exercício anterior, mesmo no caso mais geral descrito por (2.2.5) podemos dizer que se $i = 1, 2, \dots, N - 1$, π_i é uma determinada *média ponderada* de todos os demais valores π_j , $j = 0, 1, \dots, N$, $j \neq i$.
- (b) Observe que a componente mínima de π é $\pi_0 = 0$ e a componente máxima é $\pi_N = 1$.
- (c) Prove que a maior e a menor componentes do vetor π não podem aparecer também em valores “interiores” de i , ou seja, em valores $i \in \{1, 2, \dots, N - 1\}$. Você conhece propriedades análogas a esta para funções harmônicas e para o módulo de funções holomorfas (analíticas) de uma variável complexa?

3. Vamos calcular os tempos médios de absorção para o caso neutro do problema da ruína do jogador.

- (a) Mostre que no caso neutro da ruína do jogador a equação (2.3.16) fornece $\tau_1 = z_1 = N - 1$.
- (b) Usando em (2.3.15) esse resultado, mostre que $z_n = N - (2n - 1)$.
- (c) Encontre finalmente

$$\tau_n = n(N - n),$$

ou, em termos de $x = \frac{n}{N}$,

$$\tau(x) = N^2 x(1 - x).$$

Compare este resultado com (2.3.19) e (2.3.20).

Capítulo 3

Um modelo para a extinção dos Neandertais

No capítulo 1 introduzimos o conceito de cadeias de Markov e enunciámos e provámos alguns dos principais resultados. No capítulo 2 aprendemos a fazer alguns cálculos relevantes com cadeias de Markov. Estudámos em particular o processo de Moran para modelar a evolução genética em uma população. Neste último capítulo vamos usar uma cadeia de Markov de maneira importante em um modelo [?] para tentar explicar a extinção dos Neandertais e a contribuição genética que estes aportaram à humanidade.

3.1 A história natural dos homínídeos

Nós, *Homo sapiens*, somos atualmente a única espécie sobrevivente da família dos homínídeos, mas nem sempre estivemos presentes e nem sempre fomos os únicos. Vamos fazer aqui um breve resumo sobre alguns fatos da história que nos levou até a presente situação.

Os chimpanzés são a espécie viva mais parecida conosco. O número de diferenças genéticas entre as duas espécies aponta para cerca de 7 milhões de anos como uma estimativa para o tempo de divergência dos ramos da árvore evolutiva que originaram a nós e aos chimpanzés. Os primeiros fósseis que mostram diferenciação entre as linhas evolutivas do homem e do chimpanzé são datados em cerca de 6 milhões de anos. Não só os chimpanzés e gorilas, as espécies mais próximas à nossa, vivem exclusivamente na África, mas *todos* os fósseis de nossos ancestrais mais antigos foram até hoje encontrados na África. Com mais antigos queremos dizer todos os fósseis de homínídeos

datados em mais de 1 milhão de anos. Neste sentido é bastante consensual que a nossa espécie tem sua origem na África.

Fósseis de espécies homínidas nossas ancestrais datados a partir de cerca de 1 milhão de anos começam a ser encontrados fora da África. Em determinadas épocas houve coexistência de mais que uma espécie de homínido. Nessa história de família de cerca de 7 milhões de anos nossa espécie parece ser bastante recente. Os fósseis mais antigos de humanos aparentemente iguais a nós datam de cerca de 100 mil a 200 mil anos atrás e foram encontrados na Etiópia. Esses humanos *anatomicamente modernos* foram então se espalhando pelo resto do mundo, seus fósseis sendo encontrados primeiramente no Oriente Médio, depois na Ásia e na Europa e, finalmente, na Austrália. A conquista da América e da Polinésia seria muito mais tardia.

Enquanto se expandiam para fora da África os humanos anatomicamente modernos encontraram outros grupos (espécies?) homínidos descendentes dos grupos que tinham deixado a África muito antes. Em particular, existe farta documentação fóssil comprovando que os africanos anatomicamente modernos teriam existido ao mesmo tempo que os *Homo erectus* no Extremo Oriente e os Neandertais na Europa e no oeste da Ásia.

Os Neandertais foram um grupo homínido que conviveu por longo tempo com os africanos anatomicamente modernos, primeiramente no Oriente Médio e depois na Europa. Os Neandertais eram robustos, possuíam cérebros maiores que os dos humanos atuais e exibiam comportamentos avançados como uso de ferramentas sofisticadas e ornamentos corporais. Em uma caverna de Israel restos tanto de humanos anatomicamente modernos, quanto de Neandertais foram escavados em níveis diversos, documentando a ocupação alternada desse sítio por ambos os grupos por um período de cerca de 130 mil anos [? ?]. Em algum momento por volta de 25 mil anos atrás – isto é, muito, muito recentemente – os Neandertais foram misteriosamente extintos.

Não há muito o que discutir, como já comentamos há pouco, sobre a origem africana do ramo evolutivo que originou a humanidade atual. Tampouco resta dúvida de que nossa espécie deve muito de sua anatomia ao grupo africano que estamos denominando *anatomicamente moderno* e que comprovadamente se expandiu para fora da África. Aquilo que vamos discutir é se nessa expansão fora da África esse grupo originariamente africano teria se misturado geneticamente aos demais grupos homínidos que encontrou no resto do mundo, em particular com os Neandertais.

A partir dos anos 1980, com o aperfeiçoamento das tecnologias para sequenciamento do DNA, começou a surgir o que hoje chamamos a *Teoria da Origem Única Africana* (TOUA) para a humanidade. Os defensores de tal teoria afirmam que a humanidade atual ter-se-ia originado exclusivamente

do grupo africano anatomicamente moderno. Esse grupo teria convivido com os Neandertais e outros grupos primitivos, mas não teria se misturado geneticamente com eles, pois seriam *espécies distintas*. Tal teoria estava baseada inicialmente em dados genéticos, pois a comprovação fóssil só veio a aparecer mais tarde.

Os dados genéticos que davam suporte à TOUA eram oriundos principalmente do *DNA mitocondrial* de humanos vivos. Sem querer entrar aqui em mais detalhes sobre esse interessantíssimo tipo de DNA, ver por exemplo [? ? ?], um estudo amplo de sequenciamento do DNA mitocondrial feito com uma amostra de mais de uma centena de humanos vivos nos anos 1980 [?] mostrou que éramos todos virtualmente idênticos do ponto de vista do DNA mitocondrial. Ou seja, a parte algumas mutações que teriam ocorrido ao longo do tempo, tudo era compatível com toda a humanidade ser descendente de uma única mulher, que foi então chamada de *Eva mitocondrial*. Além do mais, era possível concluir que essa Eva provavelmente teria vivido na África e que o período em que viveu teria sido entre 100 mil e 200 mil anos atrás.

Segundo os defensores da TOUA, a existência de uma Eva mitocondrial africana e recente excluía a possibilidade de mistura genética dos africanos anatomicamente modernos com, por exemplo, os Neandertais. Caso essa mistura tivesse ocorrido, em vez de um único tipo básico de DNA mitocondrial para toda a humanidade, deveríamos ter pelo menos dois tipos.

A TOUA foi adotada pela quase totalidade da comunidade científica. Uma outra descoberta experimental [?] em 1997 reforçou ainda mais a teoria: conseguiu-se então sequenciar o DNA mitocondrial diretamente de um fóssil Neandertal. Descobriu-se que esse DNA era muito diferente daquele dos humanos vivos.

Ainda assim havia resistentes à TOUA e em 2005 divulgamos sobre o assunto nosso primeiro artigo de uma série de três [? ? ?]. Defendíamos então que a existência de um único tipo básico de DNA mitocondrial entre os humanos não significava que não houvera mistura por exemplo com os Neandertais, pois mesmo em um modelo de população crescente haveria a rápida extinção de linhagens de DNA mitocondrial com uma boa probabilidade de que somente uma restasse. Maurizio Serva, também físico-matemático, defendia ideias semelhantes à mesma época [? ? ?]. Como o DNA mitocondrial representa somente uma parte muito pequena dos ancestrais de um indivíduo, argumentávamos que para ter certeza sobre se os africanos anatomicamente modernos teriam se misturado ou não com outros grupos seria necessário ter mais dados sobre o DNA nuclear, que são mais difíceis de obter, de interpretar e de modelar matematicamente.

Os dados pelos quais esperávamos apareceram finalmente em 2010 [?

]: um primeiro rascunho do sequenciamento completo do DNA *nuclear* de um fóssil Neandertal. Para nossa grande sorte, a mistura genética entre Neandertais e africanos anatomicamente modernos, uma hipótese que considerávamos ainda não descartada, passava ao status de confirmada por dados experimentais. Segundo os autores do artigo, a comparação do genoma nuclear do Neandertal com os de 5 humanos vivos de várias partes do mundo – dois africanos, um europeu, um chinês e um de Papua-Nova Guiné – mostra que os não-africanos possuem genomas igual e significativamente mais próximos do genoma Neandertal do que os africanos. Estimam então que o genoma nuclear dos não-africanos atualmente vivos possui de 1 a 4% de genes de origem Neandertal. A TOUA em sua versão estrita de que não houvera mistura alguma estava derrubada.

Baseados no genoma nuclear revelado do Neandertal podemos agora afirmar com mais certeza que, como antes se supunha, a maior parte do genoma da humanidade atual vem de fato do grupo recente de africanos anatomicamente modernos que se espalhou pelo resto do mundo. Porém, ao contrário do que afirmava a TOUA, agora temos também certeza de que o grupo africano anatomicamente moderno se misturou com outros grupos humanos já presentes em outros continentes. Os Neandertais foram o primeiro grupo “primitivo” que se provou ter contribuído para a humanidade atual, mas um novo grupo – os Denisovanos – apareceu logo em estudos subsequentes [? ? ?].

3.2 Modelo para a mistura genética de duas subpopulações

O dado de que todos os não-africanos atuais são igualmente mais próximos dos Neandertais [?] sugere que a mistura entre os africanos anatomicamente modernos e os Neandertais tenha acontecido no Oriente Médio, na saída daqueles da África para o resto do mundo. A documentação fóssil, com uma caverna em Israel sendo ocupada alternadamente por Neandertais e modernos ao longo de 130 mil anos, ajuda a deixar mais crível essa hipótese [? ?].

Nesta seção vamos começar a descrever o trabalho que fizemos com Serva [?] propondo uma quantificação da mistura genética entre as subpopulações de africanos anatomicamente modernos e de Neandertais que se encontravam no Oriente Médio em algum momento do passado, antes que os modernos se disseminassem pelo resto do mundo. Além de propor um mecanismo que possa explicar a extinção subsequente dos Neandertais, temos como objetivo responder – ainda que parcialmente – às seguintes

perguntas:

1. Quão frequentes foram os episódios de mistura entre as subpopulações?
2. Quanto tempo levou a mistura até que os Neandertais fossem extintos no Oriente Médio?
3. Quantos eram os africanos e os Neandertais durante o processo de mistura?

Em vez de um modelo detalhado e com grande quantidade de parâmetros a serem ajustados, a opção feita foi a de um modelo simples e – até onde fosse possível – exatamente solúvel. O principal dado experimental a ser levado em conta é a estimativa dos autores de [?] de que 1 a 4% dos genes dos atuais não-africanos são de origem Neandertal.

Mais precisamente fazemos o seguinte conjunto de suposições:

- Os africanos anatomicamente modernos, que depois se tornaram os ancestrais dos atuais não-africanos, e uma parcela da população Neandertal conviveram em uma mesma região geográfica por um período de tempo suficientemente longo.
- Durante esse período de convivência a população total – africanos e Neandertais – permaneceu constante igual a N indivíduos.
- Ausência de obstáculos biológicos ao cruzamento. Estamos supondo explicitamente que as duas subpopulações consideradas eram indivíduos da mesma espécie.
- No entanto, provavelmente as densidades populacionais eram pequenas e os grupos eram bastante distintos do ponto de vista anatômico. É razoável supor que a mistura se tenha processado de forma lenta por causa destas razões.
- A cada geração, em média, α indivíduos de cada uma das subpopulações passava para a outra. Aqui é importante notar que o parâmetro α pode ser menor que 1 se supomos que a mistura entre as subpopulações é lenta.
- Indivíduos nascidos em uma subpopulação são considerados elementos da mesma, mesmo que um ou ambos os pais tenha vindo da outra subpopulação.
- Neutralidade: o número médio de filhos de cada indivíduo é independente da subpopulação a que pertence.

Vamos considerar o tempo discreto e, por razões que ficarão aparentes a seguir, suporemos que uma unidade de tempo corresponda a N gerações, onde N é a população total. Vamos também rotular com o índice 1 a subpopulação dos africanos e com o índice 2 a subpopulação Neandertal. A fração da população total de indivíduos da subpopulação 1 no instante t é $x(t)$, $x(t) \in [0, 1]$.

Seja $y_1(t)$ a fração de DNA nuclear africano *média* entre os indivíduos da subpopulação 1. Seja $y_2(t)$ a fração de DNA nuclear africano *média* entre os indivíduos da subpopulação 2. As equações abaixo fornecem a fração de DNA africano em cada uma das subpopulações na geração seguinte em função das frações na geração atual e levando em conta que α indivíduos da subpopulação 1 foram transferidos para subpopulação 2 e vice-versa:

$$\begin{cases} y_1(t + \frac{1}{N}) &= \left(1 - \frac{\alpha}{Nx(t)}\right) y_1(t) + \frac{\alpha}{Nx(t)} y_2(t) \\ y_2(t + \frac{1}{N}) &= \frac{\alpha}{N(1-x(t))} y_1(t) + \left(1 - \frac{\alpha}{N(1-x(t))}\right) y_2(t) \end{cases} \quad (3.2.1)$$

Nestas equações estamos usando ainda a hipótese razoável de que cada um dos α indivíduos que passam de uma subpopulação para a outra carrega consigo a fração média de DNA africano de sua subpopulação de origem. Tal *hipótese de campo médio* é bem justificada caso as trocas de indivíduos sejam tão infrequentes que a cada troca as subpopulações já tenham tido tempo de homogeneizar-se desde a troca anterior. Isto acontecerá se $\alpha \ll 1/\log_2 N$, ver [?].

Fazendo $N \rightarrow \infty$ em (3.2.1) obtemos um sistema de EDOs:

$$\begin{cases} y_1'(t) &= -\frac{\alpha}{x(t)} (y_1(t) - y_2(t)) \\ y_2'(t) &= \frac{\alpha}{1-x(t)} (y_1(t) - y_2(t)) \end{cases}, \quad (3.2.2)$$

que devem ser resolvidas usando as condições iniciais

$$y_1(0) = 1 \quad \text{e} \quad y_2(0) = 0. \quad (3.2.3)$$

Definindo variáveis auxiliares

$$z_1(t) = y_1(t) - y_2(t)$$

e

$$z_2(t) = y_1(t) + y_2(t)$$

podemos resolver o problema de valor inicial (3.2.2-3.2.3), obtendo, após cálculos simples,

$$z_1(t) = \exp \left[- \int_0^t \frac{\alpha}{x(s)(1-x(s))} ds \right] \quad (3.2.4)$$

e

$$z_2(t) = 1 + \int_0^t \frac{\alpha(2x(s) - 1)}{x(s)(1 - x(s))} z_1(s) ds . \quad (3.2.5)$$

Se conhecermos a fração $x(t) \in [0, 1]$ da subpopulação 1 na população total, essas duas fórmulas, possivelmente acompanhadas de um método numérico de integração, fornecem indiretamente as quantidades de interesse $y_1(t)$ e $y_2(t)$ a cada instante. O dado experimental de que 1 a 4% do DNA dos não-africanos atuais é de origem Neandertal significa que no instante t_e em que $x(t_e) = 1$, extinção dos Neandertais, tínhamos $y_1(t_e)$ entre 0,96 e 0,99. A partir do instante t_e o processo de mistura para porque já não temos mais duas subpopulações.

As soluções (3.2.4, 3.2.5) são facilmente interpretáveis. A primeira mostra quantitativamente que diminui com o tempo a diferença entre y_1 e y_2 , ou seja, as subpopulações vão se tornando mais parecidas uma com a outra à medida que indivíduos são trocados de um lado para o outro. Notando que a quantidade de DNA africano na subpopulação Neandertal menos a quantidade de DNA Neandertal na subpopulação Africana é $y_2 - (1 - y_1) = z_2 - 1$, a equação (3.2.5) mostra que, apesar da simetria de mesmo número de indivíduos passando de um lado para o outro, a transferência de DNA de uma subpopulação para a outra é assimétrica: quando $x > 1/2$ passa mais DNA africano para os Neandertais do que DNA Neandertal para os africanos. Mostra também que essa passagem de DNA é mais efetiva nos instantes iniciais, quando z_1 é máximo, sendo praticamente inefetiva caso z_1 chegue a ficar próximo de 0.

Até aqui, tudo é determinístico neste modelo, uma vez que a hipótese de campo médio que fizemos na dedução de (3.2.2) consiste basicamente em supor igual à média da subpopulação uma quantidade que pode variar entre os indivíduos de uma mesma subpopulação: a quantidade de DNA africano. Pergunta-se então: o que faz um modelo determinístico neste texto sobre cadeias de Markov?

A resposta será dada na próxima seção quando respondermos também a esta outra pergunta: como varia com o tempo a fração de africanos na população total? Ou seja, quais funções $x(t)$ empregar em (3.2.4, 3.2.5)?

3.3 O processo de Wright-Fisher

Respondendo diretamente as perguntas do final da última seção, as funções $x(t)$ que usaremos em (3.2.4, 3.2.5) são realizações do processo de Wright-Fisher neutro. Em particular, as $x(t)$ serão trajetórias de uma certa cadeia de Markov. Vamos dedicar esta seção para introduzir o processo

de Wright-Fisher neutro. Na próxima seção usaremos os resultados para respondermos às perguntas que já nos colocamos sobre a mistura entre os Neandertais e os ancestrais africanos dos não-africanos atuais.

Assim como o processo de Moran, que já descrevemos exaustivamente, o processo de Wright-Fisher é um modelo para a evolução temporal das frequências de genes em uma população. O processo deve seu nome a Sewall Wright e Ronald A. Fisher que, por volta dos anos 1930, fundaram com seus trabalhos independentes – junto também com John B. S. Haldane – a disciplina que hoje conhecemos como Genética de Populações. Esses três cientistas são também reconhecidos como alguns dos nomes principais da chamada *Moderna Síntese Evolutiva*: em sua época ainda não eram reconhecidas totalmente a importância da Teoria da Evolução de Darwin e da Genética de Mendel.

Assim como o processo de Moran, o de Wright-Fisher acontece em tempo discreto. Uma diferença importante é a *unidade de tempo* de cada um deles. No caso de Moran, a cada unidade de tempo acontecem exatamente um nascimento e uma morte na população. Portanto uma população de N indivíduos será bastante renovada em *cerca* de N unidades de tempo de Moran. No caso de Wright-Fisher, considera-se que toda a população é renovada ao mesmo tempo, dando lugar a uma nova *geração*. A unidade de tempo no processo de Wright-Fisher é a geração e, portanto, cada unidade de tempo de Wright-Fisher dura aproximadamente N unidades de tempo de Moran. Como os dois modelos *não são* equivalentes, não há porém como fazer uma comparação precisa entre as unidades de tempo.

Assim como no processo de Moran, no caso de Wright-Fisher trabalha-se também com as hipóteses simplificadoras de reprodução assexuada e de que o tamanho da população é fixo igual a N indivíduos. Supomos que os indivíduos da população possam ser de tipo A ou tipo B de acordo com o alelo que possuam para um determinado locus gênico. Supomos que os indivíduos de tipo A possuam aptidão relativa r e os de tipo B aptidão relativa igual a 1. Vamos nos interessar aqui principalmente pelo caso neutro $r = 1$, em que ambos os tipos têm a mesma aptidão, mas podemos por enquanto manter a generalidade.

Para completar a definição do processo de Wright-Fisher, introduzimos a função

$$\eta_r(x) = \frac{rx}{rx + 1 - x}. \quad (3.3.6)$$

Observe que η_r leva o intervalo $[0, 1]$ em si mesmo de forma sobrejetiva, se $r > 0$. No caso neutro $r = 1$, η_1 é a função identidade. Se x é a fração de indivíduos de tipo A na população em uma certa geração, então $\eta_r(x)$ é o *valor esperado* da fração de indivíduos A na próxima geração, se os

indivíduos de tipo A tiverem em média r vezes mais descendentes que os indivíduos de tipo B, ou seja, se a aptidão relativa do tipo A for r e a aptidão relativa do tipo B for 1.

Se X_n é a variável aleatória número de indivíduos de tipo A na geração n , então o processo de Wright-Fisher é definido postulando que X_{n+1} será uma variável aleatória com a distribuição binomial

$$P(X_{n+1} = j | X_n = i) = \binom{N}{j} p^j q^{N-j}, \quad (3.3.7)$$

com $p = \eta_r(i/N)$ e $q = 1 - p$.

Usando cálculos que já fizemos para a distribuição binomial, de (A.1.2) obtemos que

$$\mathbb{E}(X_{n+1}) = Np = N \frac{rX_n}{rX_n + N - X_n},$$

ou seja,

$$\mathbb{E}\left(\frac{X_{n+1}}{N}\right) = \eta_r\left(\frac{X_n}{N}\right), \quad (3.3.8)$$

conforme devíamos esperar pela definição de η_r . No caso neutro $r = 1$, o resultado acima se torna $\mathbb{E}\left(\frac{X_{n+1}}{N}\right) = \frac{X_n}{N}$, ou seja, em média as frequências dos dois tipos ficam constantes.

Como (3.3.7) satisfaz explicitamente a propriedade de Markov, o processo estocástico $(X_n)_{n=0}^{\infty}$ que acabamos de definir é claramente uma cadeia de Markov com espaço de estados $\{0, 1, \dots, N\}$, sendo o lado direito de (3.3.7) o elemento i, j da matriz de transição. Os estados 0 e N são claramente absorventes e os demais, transientes, como também no processo de Moran. Vale portanto o Teorema 1.7: com probabilidade 1 e em tempo finito a população consistirá inteiramente de um único tipo de indivíduos. O processo de Wright-Fisher seria uma espécie de versão “acelerada” do processo de Moran, realizando em uma única unidade de tempo N nascimentos e N mortes.

No caso neutro o número de indivíduos de tipo A é em média constante de uma geração para a outra, como era de se esperar. No entanto (A.1.3) mostra que a variância de X_{n+1} nunca se anula, a não ser que $X_n = 0$ ou $X_n = N$, o que causa a deriva genética no processo de Wright-Fisher. As flutuações típicas na frequência do tipo A de uma geração para a outra são

$$\frac{\sqrt{\text{Var}_{X_n}}}{N} = \frac{\sqrt{pq}}{\sqrt{N}},$$

onde usamos (A.1.3) e $p, q \in [0, 1]$ são dados acima. Esse resultado mostra mais uma vez, como comentado na Seção 2.4, que a deriva genética desaparece no limite de população infinita.

Continuamos, usando aqui os conhecimentos adquiridos no capítulo 2 para calcular as probabilidades e tempos médios de fixação para o processo de Wright-Fisher. Veremos que uma desvantagem de Wright-Fisher com relação a Moran é a impossibilidade de obter resultados exatos para essas quantidades. A razão para essa diferença é a estrutura tridiagonal simples da matriz de transição para processos de nascimento e morte, a qual não se reproduz para Wright-Fisher. Em vez de usar as formas simplificadas (2.2.7, 2.3.14) para processos de nascimento e morte, seremos obrigados a usar as formas gerais (2.2.5, 2.3.13).

Para conseguir uma maneira aproximada de calcular as probabilidades de fixação $\pi_i, i = 0, 1, \dots, N$ vamos *supor* que exista uma função *analítica* $p(x), x \in [0, 1]$ tal que $\pi_i = p(i/N)$. Fazendo $x = i/N$ a equação (2.2.5) é então reescrita como

$$p(x) = \sum_{j=0}^N t_{ji} p(x + \Delta x),$$

onde definimos também

$$\Delta x = \frac{j-i}{N}.$$

O próximo passo é *aproximar* $p(x + \Delta x)$ usando a série de Taylor de p truncada em segunda ordem:

$$p(x) \approx \sum_{j=0}^N t_{ji} [p(x) + p'(x)\Delta x + \frac{1}{2}p''(x)(\Delta x)^2].$$

Observe agora que

$$\sum_{j=0}^N t_{ji} p(x) = p(x) \sum_{j=0}^N t_{ji} = p(x),$$

onde usamos que a soma de cada coluna de uma matriz estocástica por colunas é igual a 1. Temos também

$$\begin{aligned} \sum_{j=0}^N t_{ji} p'(x)\Delta x &= \frac{p'(x)}{N} \sum_{j=0}^N t_{ji}(j-i) = \frac{p'(x)}{N} \left(\sum_{j=0}^N t_{ji} j - i \sum_{j=0}^N t_{ji} \right) \\ &= p'(x) (\eta_r(x) - x), \end{aligned}$$

onde usamos novamente a propriedade da soma de uma coluna de matriz estocástica por colunas e observamos que

$$\sum_{j=0}^N t_{ji} j = \mathbb{E}(X_{n+1}|X_n = i) = N\eta_r(i/N),$$

por (3.3.8). Finalmente,

$$\begin{aligned} \sum_{j=0}^N t_{ji} p''(x) (\Delta x)^2 &= \frac{p''(x)}{N^2} \sum_{j=0}^N t_{ji} (j-i)^2 \\ &= \frac{p''(x)}{N^2} \left(\sum_{j=0}^N t_{ji} j^2 - 2i \sum_{j=0}^N t_{ji} j + i^2 \sum_{j=0}^N t_{ji} \right), \end{aligned}$$

Das três somas acima, as duas últimas já foram calculadas. A outra é

$$\sum_{j=0}^N t_{ji} j^2 = \mathbb{E}(X_{n+1}^2 | X_n = i),$$

que calculamos usando (1.1.7) para exprimir a esperança do quadrado em termos da variância e da média e os resultados padrão de média e variância da distribuição binomial (A.1.2) e (A.1.3):

$$\sum_{j=0}^N t_{ji} j^2 = N \eta_r(x) (1 - \eta_r(x)) + N^2 \eta_r(x)^2.$$

Juntando tudo e simplificando, mostramos que p satisfaz *aproximadamente* a seguinte EDO:

$$\frac{1}{2} \left[(\eta_r(x) - x)^2 + \frac{1}{N} \eta_r(x) (1 - \eta_r(x)) \right] p''(x) + (\eta_r(x) - x) p'(x) = 0. \quad (3.3.9)$$

Lembrando que devemos acrescentar ainda as condições de contorno (2.2.4), que se escrevem aqui como

$$p(0) = 0 \quad \text{e} \quad p(1) = 1, \quad (3.3.10)$$

temos em princípio a possibilidade de calcular $p(x)$.

No caso neutro, $r = 1$ e $\eta_1(x) = x$ e (3.3.9) torna-se $p''(x) = 0$. A solução, levando em conta (3.3.10), é $p(x) = x$, ou

$$\pi_i = \frac{i}{N},$$

a mesma do caso de Moran neutro, o que não é muito surpreendente, já que os dois processos modelam de maneira diferente o mesmo fenômeno.

No caso não-neutro é possível fazer a substituição $v(x) = p'(x)$, que torna separável a EDO, de modo que as soluções podem ser escritas em

termos de integrais. Não prosseguiremos nesse caso, pois de qualquer forma as integrais resultantes não podem ser escritas em termos de funções elementares.

Podemos usar os mesmos truques para calcular *aproximadamente* os tempos médios de absorção. Substituindo em (2.3.13) $i = x/N$, $j = \frac{x+\Delta x}{N}$, supondo que $\tau_i = t(x)$ com t analítica e fazendo a mesma aproximação por um polinômio de Taylor obtemos uma EDO aproximadamente satisfeita por $t(x)$:

$$\frac{1}{2} \left[(\eta_r(x) - x)^2 + \frac{1}{N} \eta_r(x) (1 - \eta_r(x)) \right] t''(x) + (\eta_r(x) - x)t'(x) = -1 . \quad (3.3.11)$$

A solução de (3.3.11) que satisfaz as condições de contorno

$$t(0) = 0 \quad \text{e} \quad t(1) = 0 \quad (3.3.12)$$

é uma aproximação para os tempos médios de absorção no processo de Wright-Fisher. Novamente, no caso neutro é possível achar a solução explícita para $t(x)$ pois (3.3.11) torna-se

$$t''(x) = -\frac{2N}{x(1-x)} .$$

Integrando duas vezes e usando as condições de contorno, o leitor é convidado a provar que se chega ao resultado

$$t(x) \approx -2N [x \log x + (1-x) \log(1-x)] , \quad (3.3.13)$$

que é instrutivo comparar com o análogo (2.3.19) para o processo de Moran neutro. Observe que a dependência em x é idêntica, o que é bastante surpreendente levando-se em conta que ambas as fórmulas são aproximações. Por outro lado, os tempos no Moran neutro divergem com N^2 , enquanto no Wright-Fisher neutro a divergência é com N . Além do mais, existe aqui um fator multiplicativo igual a 2, inexistente no outro caso. A diferença na divergência em N é exatamente a esperada, uma que já havíamos mencionado que cada unidade de tempo em Wright-Fisher é uma geração, o que correspondente a algo da ordem de N eventos de nascimento e morte em Moran. Quanto ao fator 2 pode-se argumentar somente que os modelos não são equivalentes e portanto não é de se esperar mais correspondência entre eles do que obtivemos até aqui.

Para efeito de ilustração mostramos na Figura 3.1 algumas realizações do processo de Wright-Fisher neutro. Essas são as funções $x(t)$ típicas que precisaremos substituir em (3.2.4, 3.2.5). Obviamente não será possível

calcular exatamente as integrais necessárias. Na próxima seção falaremos mais sobre como usar o processo de Wright-Fisher neutro no modelo para a mistura entre as subpopulações de africanos anatomicamente modernos e Neandertais.

Figura 3.1: Cinco realizações do processo de Wright-Fisher neutro com $N = 100$ e $X_0 = 40$. Observe que há realizações fixando-se 0 e em N e que os tempos de fixação são da ordem de N em acordo com (3.3.13).

3.4 Aplicando o modelo de mistura

O processo de Wright-Fisher foi concebido como um modelo para a variação temporal da frequência de dois alelos A e B em uma população com reprodução assexuada. Mas a reprodução humana é sexuada, o que acarreta grandes diferenças nos modelos para frequências genéticas. Além do mais, ser africano ou Neandertal em nosso modelo não é algo que possa ser decidido por um único alelo A ou B. Pelo contrário, no modelo quantificamos quão africanos e quão Neandertais são as subpopulações através das funções $y_1(t)$ e $y_2(t)$. Um ponto que é necessário que fique claro é que o processo de Wright-Fisher neutro não está sendo usado em nosso modelo para quantificar frequências de alelos: ele está sendo introduzido somente como um mecanismo *neutro* e *natural* de flutuação nos tamanhos das subpopulações e que leve em algum tempo à extinção de uma delas. Não se trata portanto de deriva genética, mas algo que poderíamos chamar *deriva populacional*. Esta observação já responde a uma das perguntas que nos

fizemos no início da Seção 3.2: segundo o modelo aqui delineado e exposto com mais detalhes em [?], a extinção dos Neandertais teria sido *por acaso*. Em um modelo como o nosso poderia ter acontecido o contrário, com os Neandertais sobrevivendo e os Africanos sendo extintos.

O modelo que propomos para a extinção dos Neandertais possui portanto um caráter duplo. É determinístico, pois a quantidade principal a ser calculada, o valor *final* $y_1(T)$, ou seja, o valor de $y_1(t)$ no momento T da extinção dos Neandertais, aparece como solução de um sistema determinístico de EDOs. Por outro lado, também é estocástico, uma vez que dentro do sistema de EDOs aparece uma função $x(t)$ que é uma realização de um processo estocástico, o de Wright-Fisher neutro.

O objetivo do modelo é descrever sob quais condições pode ocorrer extinção dos Neandertais e, ao mesmo tempo, o valor final $y_1(T)$ cair no *intervalo experimental*, ou seja, o intervalo $0,96 < y_1(T) < 0,99$ determinado pela estimativa de [?] de que os atuais não-africanos possuem entre 1 e 4% de DNA de origem Neandertal.

Voltando atrás e recapitulando o modelo, vemos que há vários parâmetros e outras incógnitas:

- O número total N de indivíduos que participam do processo de mistura.
- A função $x(t)$, que nos conta a história detalhada da flutuação no tamanho das subpopulações até que eventualmente fossem extintos os Neandertais.
- O parâmetro α , que significa o número de indivíduos por geração que passa de cada uma das subpopulações para a outra.

Em uma observação da solução (3.2.4, 3.2.5) verifica-se que N já não está presente. O que aconteceu? O fato é que na passagem de (3.2.1) para (3.2.2) fizemos um limite quando $N \rightarrow \infty$, ocasionando o desaparecimento. Isto foi propiciado pelo fato de o tempo médio de absorção no processo de Wright-Fisher neutro ser proporcional a N e por termos escolhido a unidade de tempo como sendo N gerações. Embora N não apareça mais explicitamente na solução das EDOs (3.2.2), está escondido por exemplo em qualquer informação de tempo, por exemplo no tempo T até a extinção dos Neandertais.

A história populacional $x(t)$ é desconhecida e, conforme mostra a Figura 3.1, há uma imensa variabilidade entre os possíveis candidatos. Não só é possível que os africanos acabem extintos no lugar dos Neandertais, mas o tempo T até a extinção e como foram as oscilações da população antes que ela ocorresse são importantes. Fizemos testes substituindo várias realizações

$x(t)$ em (3.2.4, 3.2.5) e calculamos numericamente as integrais ou, o que é equivalente, substituímos diretamente as realizações em (3.2.2) e usamos um algoritmo numérico para aproximar a solução. O que se vê é que a forma do gráfico de $x(t)$ influencia fortemente o valor de $y_1(t)$ no momento da extinção dos Neandertais. Por outro lado, como N desapareceu, para gerar histórias $x(t)$ o único parâmetro necessário, e que influencia fortemente o valor de T , é $x(0)$, o valor inicial, na época do primeiro contato entre as duas subpopulações, da fração de africanos na população.

Uma vez que α mede a “intensidade” do processo de mistura, seu valor terá também enorme influência no valor final de $y_1(t)$. Para uma mesma história $x(t)$ o valor final $y_1(T)$ pode ser tão próximo a 1 quanto se queira, desde que α seja pequeno o suficiente, ou pequeno demais se α for grande.

Com tantas incógnitas, a solução que encontramos foi:

1. Sortear um valor para $x(0)$ em $[0, 1]$ e gerar uma realização $x(t)$ do processo de Wright-Fisher neutro.
2. Descartar a realização, caso os africanos tenham sido extintos.
3. Caso os Neandertais tenham sido extintos, sortear um valor de α e calcular numericamente o valor final $y_1(T)$.
4. Caso $y_1(T)$ caia no intervalo experimental, anotam-se os valores de α , $x(0)$ e T . Caso contrário, descarta-se a realização.

Repetimos – ou melhor, quem repetiu foi o computador – muitas vezes os passos acima até que tivéssemos acumulado um milhão de eventos de sucesso, ou seja, eventos em que os Neandertais foram extintos e $y_1(T)$ caiu no intervalo experimental. Com os dados desses eventos contamos quantos deles caíam em pequenos intervalos de α , construindo o histograma da Figura 3.2. Do histograma vê-se que o valor mais provável para α , a abscissa do ponto de máximo, é $\alpha_{max} \approx 0,013$, que corresponde a um par de indivíduos trocados entre as duas subpopulações a cada 77 gerações. Por outro lado, valores de α maiores que α_{max} ocorrem com grande probabilidade, de forma que o valor médio de α no universo dos eventos mostrados na Figura 3.2 é $\alpha_{medio} \approx 0,083$, correspondendo a um par de indivíduos trocados a cada 12 gerações. Estas estimativas respondem à primeira das perguntas do início da Seção 3.2.

Para responder às outras duas perguntas produzimos uma amostra menor com somente 790 eventos de sucesso. Para cada evento da amostra colocamos em um gráfico o valor de $x(0)$ na abscissa e o valor do tempo de extinção T na ordenada. O resultado está mostrado na parte da esquerda da Figura 3.3. Na parte da direita da mesma figura colocamos, para a mesma amostra de 790 eventos, $x(0)$ na abscissa e α na ordenada.

Figura 3.2: A densidade de probabilidade para o parâmetro α , a taxa de mistura entre as populações, obtida de maneira empírica através da simulação de um milhão de eventos com extinção dos Neandertais e valor final $y_1(T)$ no intervalo experimental entre 96% e 99%. Dessa figura, o valor mais provável é $\alpha_{max} \approx 0,013$, enquanto a média é $\alpha_{medio} \approx 0,083$.

Figura 3.3: Usando uma mesma amostra de 790 eventos com extinção dos Neandertais e valor final $y_1(T)$ no intervalo experimental entre 96% e 99% mostramos à esquerda a correlação entre a fração inicial $x(0)$ de africanos na população e o tempo de extinção T medido em gerações divididas por N e à direita a correlação entre $x(0)$ e a taxa de mistura α . O tempo médio de extinção na amostra foi de aproximadamente 0,58 e o desvio padrão de aproximadamente 0,55.

Como mostra o lado direito da Figura 3.3, é impossível ser muito preciso com relação ao tempo que transcorreu até a extinção dos Neandertais, uma vez que a variabilidade nos tempos de extinção é bastante grande. Tomando a média da amostra como uma ideia da ordem de grandeza dos tempos envolvidos, temos $T \approx 0,58$, lembrando que uma unidade de tempo significa N gerações. Tomando o valor usual de 20 anos como duração de uma geração humana e supondo que o tempo de extinção foi de 130000 anos, como sugerido pelo dado em [?], obtemos uma estimativa para o tamanho da população: $N \approx 11000$ indivíduos, valor bastante compatível com estimativas da população humana em eras pré-históricas, ver por exemplo [?].

O lado esquerdo da mesma figura mostra que quanto maior o valor de $x(0)$ menos preditivo o modelo é para α . Existem razões matemáticas para tal, mas remetemos o leitor interessado ao artigo [?] para mais detalhes. Mostra também que valores maiores de α são incompatíveis com valores pequenos de $x(0)$. Vemos ainda que do ponto de vista do modelo também é de esperar que os africanos fossem a maioria da população no instante inicial.

Encerramos este texto mencionando que em outro artigo [?] analisamos a possibilidade de que os africanos tivessem aptidão maior que a dos Neandertais. O modelo é basicamente o mesmo tratado aqui, usando-se porém o processo de Wright-Fisher não-neutro. Há algumas diferenças técnicas interessantes, mas uma consequência da não-neutralidade é que o uso do formalismo estocástico perde importância e uma aproximação determinística foi capaz de produzir bons resultados. Outra consequência da perda da neutralidade é que os tempos de extinção são substancialmente menores. Concluímos em [?] que mesmo para diferenças de aptidão tão pequenas quanto 1% o modelo não-neutro é incapaz de explicar a convivência alternada de africanos e Neandertais em uma mesma caverna por períodos tão longos quanto os 130000 anos reportados em [? ?].

Exercícios do capítulo 3

- Um caso interessante onde é possível resolver exatamente o problema de valor inicial (3.2.2, 3.2.3) ou, equivalentemente, as integrais em (3.2.4, 3.2.5) é quando as frações das subpopulações são constantes: $x(t) = c$. Sejam $\phi_1(c, t)$ e $\phi_2(c, t)$ as soluções respectivamente para y_1 e y_2 no caso em que $x(t) = c$.
 - Obtenha fórmulas explícitas para $\phi_1(c, t)$ e $\phi_2(c, t)$.

-
- (b) O que acontece com $\phi_1(c, t)$ e $\phi_2(c, t)$ quando $t \rightarrow \infty$? É este o comportamento esperado? Por quê?
- (c) Prove que se existem constantes c_1 e c_2 tais que $x(t) \in [c_1, c_2]$ para $t \in [0, T]$, então as soluções do problema de valor inicial (3.2.2, 3.2.3) satisfazem $\phi(c_1, t) \leq y_2(t) \leq y_1(t) \leq \phi(c_2, t)$ se $t \in [0, T]$.
2. Vamos analisar um outro caso interessante, mas agora não é possível solução exata do problema de valor inicial (3.2.2, 3.2.3). Vamos supor $x(t) = 0,7 - 0,2 \cos(5t)$ com $\alpha = 1$ e $t \in [0, 2]$.
- (a) Use um programa como *Mathematica*, *Maple* ou *Matlab* para obter a solução numérica do problema de valor inicial. Faça um gráfico conjunto de y_1 e y_2 .
- (b) Encontre valores de c_1 e c_2 tais que $x(t) \in [c_1, c_2]$ para $t \in [0, 2]$.
- (c) Faça um gráfico mostrando que o resultado do item (c) do problema anterior foi verificado.

Apêndice A

Distribuições binomial e geométrica

Neste apêndice definiremos duas importantes distribuições de probabilidade para variáveis aleatórias discretas e calcularemos a média e a variância para as variáveis aleatórias que seguem essas distribuições. Como diversos problemas levam a essas distribuições, os cálculos que aqui faremos são de grande e geral utilidade. O cálculo de propriedades das várias distribuições de probabilidade é assunto padrão que pode ser encontrado nos livros de Probabilidade, por exemplo [? ? ?].

A.1 Distribuição binomial

Vimos no Exemplo 1.4 o aparecimento de uma variável aleatória que é a soma de N variáveis independentes, cada uma das quais representando o aparecimento de cara em uma das moedas. Se definirmos como *sucesso* o aparecimento de uma cara, então, em geral, a distribuição binomial é aquela que se aplica para calcular o número médio de sucessos em N repetições independentes de um mesmo fenômeno, onde as probabilidades de sucesso p e de fracasso $q = 1 - p$ são constantes.

Definimos a distribuição binomial com parâmetros de sucesso $p \in [0, 1]$ e fracasso $q = 1 - p$ como sendo a função de probabilidade f de uma variável aleatória X que pode assumir os valores $0, 1, 2, \dots, N$, $N \in \mathbb{N}$, onde

$$P(X = n) = f(n) = \binom{N}{n} p^n q^{N-n}, \quad (\text{A.1.1})$$

ver (1.1.9).

O nome *binomial* vem do fato de que, assim como já fizemos logo após (1.1.9), a fórmula do binômio de Newton será usada repetidamente em nossos cálculos.

Se X é binomialmente distribuída, sua média será então

$$\begin{aligned}\mathbb{E}(X) &= \sum_{n=0}^N n f(n) = \sum_{n=1}^N n \binom{N}{n} p^n q^{N-n} \\ &= p \frac{\partial}{\partial p} \sum_{n=0}^N \binom{N}{n} p^n q^{N-n},\end{aligned}$$

onde, na última igualdade, fizemos aparecer uma derivada parcial supondo que p e q sejam variáveis independentes. Identificando agora a soma na última expressão como a fórmula do binômio de Newton para $(p+q)^N$ continuamos com

$$\begin{aligned}\mathbb{E}(X) &= p \frac{\partial}{\partial p} (p+q)^N = p N (p+q)^{N-1} \\ &= Np,\end{aligned}$$

onde usamos finalmente, no último passo, que $p+q=1$. Portanto, como usado em (1.1.10),

$$\mu_X = Np. \quad (\text{A.1.2})$$

Para calcular Var_X , começamos com

$$\begin{aligned}\mathbb{E}(X^2 - X) &= \sum_{n=2}^N n(n-1) \binom{N}{n} p^n q^{N-n} = p^2 \frac{\partial^2}{\partial p^2} \sum_{n=0}^N \binom{N}{n} p^n q^{N-n} \\ &= p^2 N(N-1)(p+q)^{N-2} = N(N-1)p^2.\end{aligned}$$

Então

$$\mathbb{E}(X^2) = \mathbb{E}(X^2 - X) + \mathbb{E}(X) = N(N-1)p^2 + Np,$$

onde usamos (A.1.2). Finalmente, usando (1.1.7), (A.1.2) de novo, e fáceis simplificações, chegamos a

$$\text{Var}_X = Npq. \quad (\text{A.1.3})$$

A.2 Distribuição geométrica

Analogamente ao que fizemos na Seção A.1, começemos com um modelo real para motivar uma distribuição de probabilidades. Aqui vamos nos

perguntar quantas vezes em média devemos lançar um dado até que este caia em 6. Ou, de forma parecida, quantas vezes em média devemos jogar em uma loteria até que ganhemos o prêmio. A ideia geral é que fazemos diversas tentativas *independentes* para tentar obter um evento de “sucesso”, onde o sucesso pode acontecer em qualquer das tentativas (mas é em geral um evento pouco provável, mas possível). No caso do lançamento dos dados, o sucesso é o aparecimento do 6. No caso da loteria, o sucesso é ganhar o prêmio.

Suponha então que $p \in (0, 1]$ é a probabilidade de sucesso em cada tentativa. Vamos também denotar por $q = 1 - p$ a probabilidade de fracasso em cada tentativa. Seja X a variável aleatória que indica o número da tentativa em que o sucesso aconteceu pela primeira vez. O espaço de estados de X é portanto $\{1, 2, \dots\}$ e podemos nos perguntar pela probabilidade de cada um dos valores de X .

O evento $\{X = 1\}$ significa que na primeira tentativa tivemos sorte e o sucesso aconteceu. Se p é a probabilidade de sucesso em cada tentativa, então obviamente

$$P(X = 1) = p.$$

O evento $\{X = 2\}$ significa que a primeira tentativa falhou, mas o sucesso foi alcançado na segunda. Como as tentativas são, por hipótese, independentes, a probabilidade de um fracasso seguido por um sucesso é

$$P(X = 2) = qp.$$

Seguindo o raciocínio aqui sugerido, então a probabilidade de que o primeiro sucesso aconteça somente na tentativa n , $n = 1, 2, \dots$, é a probabilidade de $n - 1$ fracassos seguidos por um sucesso:

$$P(X = n) = q^{n-1}p.$$

A distribuição geométrica com parâmetro de sucesso $p \in (0, 1]$ é definida como a distribuição de uma variável aleatória X em que

$$P(X = n) = q^{n-1}p. \quad (\text{A.2.4})$$

O nome distribuição geométrica vem do fato de que a sequência dos valores $P(X = n)$, $n = 1, 2, \dots$ é uma *progressão geométrica* de razão $q \in [0, 1)$.

O leitor que joga na loteria e até hoje nunca ganhou pode ter alguma esperança, pois, como a razão da progressão geométrica é estritamente menor que 1, então a probabilidade de que tenha que esperar n jogadas até ganhar pela primeira vez tende a 0 quando $n \rightarrow \infty$. Por outro lado, seria interessante saber o tempo médio que terá que esperar até ganhar pela primeira vez. É o que faremos agora.

Antes disto, para aquecer os motores, calculemos

$$\sum_{n=1}^{\infty} P(X = n) .$$

Se tudo estiver correto em nossos raciocínios, por (iii) e (iv) na Definição 1.1 devemos ter essa soma igual a 1. Conferindo,

$$\sum_{n=1}^{\infty} P(X = n) = \sum_{n=1}^{\infty} q^{n-1} p = p \sum_{n=1}^{\infty} q^{n-1} .$$

Mas a soma da progressão geométrica infinita converge a $\frac{1}{1-q}$ se a razão q satisfizer $|q| < 1$. Em particular se $q \in [0, 1)$, como aqui. Logo

$$\sum_{n=1}^{\infty} P(X = n) = p \frac{1}{1-q} = 1 ,$$

pois $q = 1 - p$.

O cálculo de $\mathbb{E}(X)$ prossegue agora na mesma linha:

$$\begin{aligned} \mathbb{E}(X) &= \sum_{n=1}^{\infty} n P(X = n) = p \sum_{n=1}^{\infty} n q^{n-1} = p \frac{d}{dq} \sum_{n=1}^{\infty} q^n \\ &= p \frac{d}{dq} \frac{q}{1-q} , \end{aligned}$$

onde usamos novamente o resultado da soma de uma progressão geométrica infinita. Fazendo um cálculo simples de derivada e algumas simplificações, chegamos finalmente ao resultado

$$\mathbb{E}(X) = \frac{1}{p} . \tag{A.2.5}$$

Convido ao leitor que calcule o tempo médio que terá que esperar até ganhar na loteria e a probabilidade de que ganhe pelo menos uma vez durante sua vida. É possível também calcular Var_X , que fica como exercício.

Bibliografia

- [1] L. J. S. Allen. *An Introduction to Stochastic Processes with Applications to Biology*. Pearson, New Jersey, 2003.
- [2] G. Ávila. A série harmônica e a fórmula de Euler-MacLaurin. *Matemática Universitária*, 19:55, 1995.
- [3] O. Bar-Yosef. Neandertals and modern humans in western Asia. In T. Akazawa, K. Aoki, and O. Bar-Yosef, editors, *The Chronology of the Middle Paleolithic of the Levant*, pages 39–56. Plenum, New York, 1999.
- [4] R.G. Bartle. *The Elements of Integration and Lebesgue Measure*. Wiley Classics Library. Wiley, 2011.
- [5] R. L. Cann, M. Stoneking, and A. C. Wilson. Mitochondrial DNA and human evolution. *Nature*, 325:31, 1987.
- [6] Bernard Derrida, Susanna C. Manrubia, and Damián H. Zanette. Statistical properties of genealogical trees. *Phys. Rev. Lett.*, 82(9):1987–1990, Mar 1999.
- [7] L. Edelstein-Keshet. *Mathematical Models in Biology*. Society for Industrial and Applied Mathematics, Philadelphia, 2005.
- [8] W. Feller. *An Introduction to Probability Theory and Its Applications*. Wiley, 1960.
- [9] Pedro J. Fernandez. *Introdução à Teoria das Probabilidades*. Editora UnB, 1973.
- [10] Pedro J. Fernandez. *Medida e Integração*. Projeto Euclides. IMPA, Rio de Janeiro, 1976.

-
- [11] F. Gantmacher. *The theory of matrices, vol. 2*. Chelsea Publishing Company, 1959.
- [12] Richard E. Green, Johannes Krause, and et al. A draft sequence of the Neandertal genome. *Science*, 328:710–722, 2010.
- [13] Ben Harder. Did Humans and Neandertals battle for control of the Middle East?. Disponível em http://news.nationalgeographic.com/news/2002/03/0305_0307_neandertal.html. Acessado em 17/09/2012. *National Geographic News*, March 8, 2002.
- [14] M. Krings, A. Stone, R. W. Schmitz, H. Krainitzki, M. Stoneking, and S. Paabo. Neandertal DNA sequences and the origin of modern humans. *Cell*, 90:19, 1997.
- [15] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2009.
- [16] L. Martignon. *Matrizes positivas*. 16º Colóquio Brasileiro de Matemática. Instituto de Matemática Pura e Aplicada, 1987.
- [17] H. Y. Minc. *Nonnegative Matrices*. John Wiley and Sons, 1988.
- [18] J. D. Murray. *Mathematical Biology I. An Introduction, 3rd edition*. Springer, New York, 2002.
- [19] Armando G. M. Neves. Interbreeding conditions for explaining Neandertal DNA in living humans: the nonneutral case. In R. P. Mondaini and R. Dilão, editors, *BIOMAT 2011*. World Scientific, 2012.
- [20] Armando G. M. Neves and Carlos H. C. Moreira. The mitochondrial Eve in an exponentially growing population and a critique to the Out of Africa model for human evolution. In R. P. Mondaini and R. Dilão, editors, *BIOMAT 2005*. World Scientific, 2005.
- [21] Armando G. M. Neves and Carlos H. C. Moreira. Applications of the Galton-Watson process to human DNA evolution and demography. *Physica A*, 368:132, 2006.
- [22] Armando G. M. Neves and Carlos H. C. Moreira. The number of generations between branching events in a Galton-Watson tree and its application to human mitochondrial DNA evolution. In R. P. Mondaini and R. Dilão, editors, *BIOMAT 2006*. World Scientific, 2006.

-
- [23] Armando G. M. Neves and Maurizio Serva. Extremely rare interbreeding events can explain Neanderthal DNA in modern humans. *PLoS ONE*, 7(10):e47076, 2012.
- [24] Morten Rasmussen, Xiaosen Guo, and et al. An aboriginal Australian genome reveals separate human dispersals into Asia. *Science*, 334(6052):94–98, 2011.
- [25] D. Reich, Richard E. Green, and et al. Genetic history of an archaic hominin group from Denisova cave in Siberia. *Nature*, 468(7327):1053–1060, 2010.
- [26] David Reich, Nick Patterson, and et al. Denisova admixture and the first modern human dispersals into southeast Asia and Oceania. *The American Journal of Human Genetics*, 89(4):516 – 528, 2011.
- [27] S. Ross. *Probabilidade: Um Curso Moderno com Aplicações*. Bookman, 2010.
- [28] Maurizio Serva. Lack of self-averaging and family trees. *Physica A*, 332:387 – 393, 2004.
- [29] Maurizio Serva. On the genealogy of populations: trees, branches and offspring. *J. Stat. Mech.-Theory E*, 2005(07):P07011, 2005.
- [30] Maurizio Serva. Mitochondrial DNA replacement versus nuclear DNA persistence. *J. Stat. Mech.-Theory E*, 2006(10):P10013, 2006.
- [31] N. Takahata. Allelic genealogy and human evolution. *Mol. Biol. Evol.*, 10(1):2, 1993.

Índice

- Alelos, 27
- Aptidão, 54, 76
- Autovalor dominante, 38, 40
- Autovetor, 37

- Cadeia de Markov, 24
- Cadeia de Markov aperiódica, 42
- Cadeia de Markov ergódica, 42
- Cadeia de Markov periódica, 42, 48
- Caminho, 35
- Classes de comunicação, 44
- Classes de equivalência, 44
- Comprimento, 35, 44, 49
- Constante de Euler-Mascheroni, 63

- Deriva genética, 13, 30, 65, 77
- Digrafo, 35
- Digrafo fortemente conexo, 36
- Disco de Gershgorin, 39
- Distribuição binomial, 23, 77
- Distribuição de equilíbrio, 34, 41
- Distribuição geométrica, 46
- DNA mitocondrial, 71

- Espaço amostral, 17
- Esperança, 21
- Estado absorvente, 44
- Estado acessível, 43
- Estado recorrente, 45
- Estado transiente, 45
- Estados comunicantes, 44
- Eventos, 17
- Eventos independentes, 19

- Fixação, 28, 29, 47
- Função de probabilidade, 20

- Grafo, 35

- Irredutibilidade, 36

- Lançamento de N moedas, 20, 22
- Lançamento de dois dados, 18, 19

- Média de uma variável aleatória, 21
- Matriz de transição, 30
- Matriz estocástica, 31, 40, 78
- Matriz irredutível, 36, 38
- Matriz não-negativa, 35
- Medida, 18
- Mutação, 27

- Neandertais, 70

- Perron-Frobenius, 35, 36
- Princípios do máximo e do mínimo, 67
- Probabilidade, 17
- Probabilidade condicional, 19
- Probabilidades de fixação, 55, 78
- Probabilidades de primeiro retorno, 44
- Processo de Moran, 27, 35, 36, 45, 51, 76
- Processo estocástico, 24
- Processos de nascimento e morte, 14, 51, 52, 56, 62

-
- Relação de equivalência, 44
Ruína do jogador, 14, 51, 53, 58, 64
- Série harmônica, 63
SIR, Modelo, 25
- Tempo médio de fixação, 60
Teorema de Gershgorin, 39
Teorema de Perron-Frobenius, 14, 38
Teorema de Perron-Frobenius para
matrizes irredutíveis, 38
Teoria da Origem Única Africana, 70
Traço de uma matriz, 39
- Variáveis aleatórias independentes, 21
Variável aleatória, 20
Variância de uma variável aleatória,
22