Capítulo 3

Probabilidade

Neste capítulo introduzimos modelos probabilísticos como modelos matemáticos para experimentos aleatórios. A aplicação principal, mas não a única, a ser discutida ao longo deste capítulo e do próximo, será à amostragem (de variáveis) em populações, objeto básico da *inferência estatística*.

Inferência estatística

Um dos problemas práticos na análise de distribuições de variáveis em populações, como fizemos nos primeiros capítulos, é obter os dados populacionais que nos permitam determinar as distribuições, ou suas características. É preciso observar *toda* a população, e isto pode ser muito custoso.

A alternativa é examinar apenas uma *amostra* da população, uma parte desta. A questão então é como podemos extrapolar, *inferir* da informação amostral as características populacionais. Surge a questão de representatividade da amostra: é preciso garanti-la de alguma forma. Na falta de melhor opção (que é o que costuma acontecer), uma forma fraca de fazê-lo é tomar amostras aleatórias, ou casuais, ou sorteadas. No caso mais simples, que é o que consideraremos adiante, não temos mais representatividade absoluta, em que cada indivíduo da população é representado, mas uma representatividade *probabilística*: cada indivíduo da população tem *a mesma chance* de ser representado.

As inferências que fazemos a partir de amostras aleatórias não têm validade absoluta, mas apenas uma validade probabilística. Atribuir probabilidades a tais inferências envolve fazermos uma descrição detalhada das probabilidades envolvidas nos sorteios, e sabermos fazer cálculos com elas. É o que aprenderemos neste capítulo, desde um ponto de vista mais geral.

3.1 Modelo probabilístico

Em diversas situações, na natureza, na sociedade, ocorrem fenômenos que podemos chamar de aleatórios: são aqueles que, mesmo quando observados repetidamente sob as mesmas condições, produzem resultados diferentes, de forma *imprevisível*. Tais fenômenos podem ser vistos como o resultado de *experimentos aleatórios*.

Exemplos

Jogos de azar, como o lançamento de um dado, ou uma rodada de roleta, ou a distribuição de mãos de carteado, têm resultados imprevisíveis, e podem ser então considerados como *experimentos aleatórios*.

Fenômenos práticos, como o tempo que fará no final de semana, o resultado de um evento esportivo, o rendimento de uma carteira de investimentos, medições com instrumentos, também podem ser considerados como resultado de experimentos aleatórios em muitas situações. O resultado de uma amostragem aleatória em uma população seria um outro exemplo.

A natureza fornece por sua vez muitos exemplos: características diversas de animais e plantas, comportamentos climáticos, sísmicos, marítimos, cósmicos, etc.

Apesar da imprevisibilidade sobre o resultado de sua próxima observação, muitos destes experimentos apresentam uma regularidade ou previsibilidade *estatística*: ao longo de muitas observações repetidas do experimento, as freqüências relativas dos diversos resultados possíveis se estabilizam.

Exemplo 3.1 *No lançamento de um dado equilibrado, observa-se que as freqüências relativas de* $1, 2, 3, 4, 5$ *ou* 6 *se aproximam de* $1/6$ *cada conforme o número de repetições vai aumentando. Na Tabela 3.1 indicamos tais freqüências relativas para um número crescente de repetições.*

Por isto, ou por outra motivação intuitiva, atribuímos *probabilidades* aos diversos resultados possíveis de (certos) experimentos aleatórios. No caso do lançamento de um dado equilibrado, é natural atribuir probabilidade $1/6$ a cada um dos 6 possíveis resultados.

Modelo probabilístico

Num modelo matemático para um experimento aleatório, vamos abstrair os ingredientes essenciais. Um deles é a *multiplicidade* de resultados possíveis.

Resultado	freqüência		
1	0.180	0.170	0.163
2	0.180	0.171	0.166
3	0.200	0.164	0.174
4	0.130	0.148	0.162
5	0.130	0.175	0.170
6	0.180	0.172	0.166
N	100	1000	10000

Tabela 3.1 N lançamentos de um dado equilibrado (simulação)

Isto será indicado por um conjunto Ω não vazio (e tipicamente não unitário), que chamaremos de *espaço amostral*.

No Exemplo 3.1 teríamos

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

O outro ingrediente são as probabilidades, que em princípio devem ser atribuídas aos resultados possíveis (na linguagem do modelo, aos pontos de Ω). No mesmo exemplo, como o dado é equilibrado, nenhum valor teria mais chance de sair do que os outros, e teríamos

$$\mathbb{P}(1) = \mathbb{P}(2) = \mathbb{P}(3) = \mathbb{P}(4) = \mathbb{P}(5) = \mathbb{P}(6) = \frac{1}{6}, \tag{3.1}$$

onde $\mathbb{P}(i)$ lê-se "probabilidade de (sair) i (como resultado do lançamento do dado)".

De forma mais geral, pode ocorrer de uma atribuição de probabilidades aos *pontos* de Ω (como fizemos acima) não fazer muito sentido ou não ser o suficiente.

Exemplo 3.2 *Um experimento aleatório razoavelmente familiar (bastante, para quem faz simulações) é: escolha ao acaso e de maneira uniforme um número do intervalo* $[0, 1]$*. A uniformidade implica que* de certa forma *cada resultado possível deve ter a mesma probabilidade. Mas não podemos atribuir probabilidade igual a todos os pontos de* $\Omega = [0, 1]$ *(há um infinito contínuo de possibilidades), sem que essa probabilidade seja zero. E mesmo fazendo isto, esta atribuição é insuficiente.*

A maneira adequada de fazer a atribuição neste caso é, por exemplo, atribuir probabilidades aos intervalos. Da uniformidade, seria natural impor *que para todo subintervalo I de $[0,1]$,*

$$\mathbb{P}(I) = \textit{comprimento de } I,$$

onde $\mathbb{P}(I)$ significa probabilidade de (o número escolhido pertencer a) I. (Note que neste caso a probabilidade de um ponto é igual ao seu comprimento, que se anula.)

Note ainda no exemplo acima que não basta atribuir probabilidades aos pontos de Ω: é necessário considerarmos subconjuntos adequados, no caso, os intervalos.

Genericamente então, num modelo probabilístico, as probabilidades são atribuidas a subconjuntos de Ω, os *eventos.*

Eventos

Dado um modelo probabilístico (para certo experimento aleatório) com espaço amostral (conjunto de possibilidades) Ω, e um subconjunto $A \subset \Omega$, dizemos que A é um *evento*, e que, no contexto do experimento aleatório, A *ocorre* se o resultado do experimento (um ponto de Ω) pertencer a A.

No Exemplo 3.2, $A = [0, 1/2]$ é um evento, que ocorre se o número escolhido for menor ou igual a $1/2$.

Espaço de eventos

O espaço de eventos do modelo probabilístico, que podemos denotar por $\mathcal{E}$, é o conjunto (ou classe) de eventos que queremos considerar (e atribuir probabilidades). Vamos definir operações (entre eventos) nesta classe, que deverá ser rica o suficiente para ser preservada pelas operações (isto é, quando aplicarmos as operações a eventos da classe, o resultado deve ser um evento da classe). As operações são:

1. com dois eventos: intersecção (ocorrência simultânea); união (ocorrência alternativa);

2. com um evento: complementação (não ocorrência).

Dados dois eventos $A, B \in \mathcal{E}$, a intersecção

$$A \cap B \in \mathcal{E}$$

é um evento (da classe $\mathcal{E}$): a ocorrência simultânea de A *e* B; e a união

$$A \cup B \in \mathcal{E}$$

é um evento (da classe $\mathcal{E}$): a ocorrência alternativa de A *ou* B (ocorre A ou B — incluindo a possibilidade de ambos ocorrerem).

Seja ainda a *diferença* entre A e B:

$$A \setminus B \in \mathcal{E},$$

que é um evento (da classe $\mathcal{E}$): ocorre A mas não ocorre B.

Dado um evento $A \in \mathcal{E}$, o complemetar de A

$$A^c \in \mathcal{E}$$

é um evento (da classe $\mathcal{E}$): a não ocorrência de A.

Observação 3.3 *As operações acima podem e devem ser pensadas como as operações usuais entre conjuntos com a mesma terminologia e notação. (Em termos de conjuntos, $A \backslash B$ é o conjunto de elementos de A que não pertencem a B.)*

Valem as seguintes relações (verifique). Dados $A, B \in \mathcal{E}$

$$\begin{aligned} (A^c)^c &= A & (3.2)\\ A \setminus B &= A \cap B^c & (3.3)\\ (A \cap B)^c &= A^c \cup B^c & (3.4)\\ (A \cup B)^c &= A^c \cap B^c. & (3.5) \end{aligned}$$

As propriedades (3.4) e (3.5) são prontamente generalizáveis para $n \geq 2$ eventos. Sejam $A_1, A_2, \ldots, A_n \in \mathcal{E}$ tais eventos. Então

$$\begin{aligned} \left(\cap_{i=1}^n A_i\right)^c &= \cup_{i=1}^n A_i^c, & (3.6)\\ \left(\cup_{i=1}^n A_i\right)^c &= \cap_{i=1}^n A_i^c. & (3.7) \end{aligned}$$

Os subconjuntos $\emptyset$ e Ω de Ω também são eventos: o evento nulo ou impossível, e o evento total ou certo, respectivamente. Eles sempre estarão em $\mathcal{E}$.

Dizemos que dois eventos $A, B \in \mathcal{E}$ são *disjuntos* ou *mutuamente exclusivos* se

$$A \cap B = \emptyset. \tag{3.8}$$

No Exemplo 3.1, uma classe de eventos natural é o *conjunto das partes* de $\Omega = \{1, 2, 3, 4, 5, 6\}$, isto é, todos os subconjuntos de $\{1, 2, 3, 4, 5, 6\}$:

$$\begin{aligned}\mathcal{E} &= \{\emptyset, \{1\}, \{2\}, \ldots, \{6\}, \{1, 2\}, \ldots, \{1, 6\}, \\ &\quad \{2, 3\}, \ldots, \{2, 6\}, \ldots, \{6, 6\}, \{1, 2, 3\}, \ldots, \Omega\}.\end{aligned}$$

Note que, por exemplo, $\{1, 2, 3\}$ e $\{4, 6\}$ são eventos disjuntos de $\mathcal{E}$.

Não é difícil verificar que o número de elementos de $\mathcal{E}$, neste exemplo, é $2^6 = 64$: dado um evento A de $\mathcal{E}$, para cada ponto i de Ω, seja

$$X_i = \begin{cases} 1, & \text{se } i \in A, \\ 0, & \text{se } i \notin A. \end{cases}$$

Então A é determinado pela (além de determinar a) seqüência $X_1, X_2, \ldots, X_6$. Logo há uma relação 1 a 1 entre os eventos de $\mathcal{E}$ e as seqüências de 0's e 1's de comprimento 6. Como há duas possibilidades para cada uma das 6 entradas da seqüência (0 ou 1), o total de possibilidades para a seqüência é $2 \times 2 \times \ldots 2 = 2^6$.

Uma outra classe que poderia ser considerada, caso estivéssemos interessados apenas se o lançamento resulta em número par ou ímpar, seria

$$\mathcal{E}' = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}.$$

Note que esta classe de eventos tem apenas 4 elementos.

No Exemplo 3.2, uma classe de eventos adequada não é simples de descrever; em particular o conjunto das partes de $[0, 1]$ é grande demais. Mas não nos preocuparemos com isto nestas notas, e apenas diremos que neste caso $\mathcal{E}$ deve conter todos os intervalos de Ω (incluindo os de comprimento 0: os conjuntos unitários e o conjunto vazio) e uniões de tais intervalos.

Probabilidade

Dados um espaço amostral Ω e uma classe de eventos $\mathcal{E}$, uma *probabilidade* é uma função $\mathbb{P} : \mathcal{E} \to [0, 1]$ satisfazendo as seguintes propriedades.

$$\mathbb{P}(\Omega) = 1 \quad \text{(o evento certo tem probabilidade 1)} \tag{3.9}$$

Dados dois eventos A e B *disjuntos* (veja (3.8))

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \quad \text{(aditividade).} \tag{3.10}$$

A imposição destas duas propriedades é intuitivamente clara. Note que as distribuições de freqüência vistas nos primeiros capítulos têm propriedades semelhantes (veja (1.7), (1.14), (1.18)).

Definição 3.4 *A tripla* $(\Omega, \mathcal{E}, \mathbb{P})$ *é o que chamaremos de* modelo probabilístico *ou* espaço de probabilidades *(para dado experimento aleatório).*

De (3.9) e (3.10) podemos deduzir uma série de outras propriedades da probabilidade.

1. Complementaridade: dado $A \in \mathcal{E}$
$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A). \tag{3.11}$$

2. $$\mathbb{P}(\emptyset) = 0 \quad \text{(o evento nulo tem probabilidade 0)} \tag{3.12}$$

3. Dados $A, B \in \mathcal{E}$ tais que $A \subset B$, então
$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A). \tag{3.13}$$

4. Regra da soma: dados $A, B \in \mathcal{E}$ *quaisquer*
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B). \tag{3.14}$$

Demonstrações

Como A e A^c são disjuntos e $A \cup A^c = \Omega$, de (3.9) e (3.10),
$$\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(A \cup A^c) = \mathbb{P}(\Omega) = 1, \tag{3.15}$$
e (3.11) segue.

Tomando $A = \Omega$ em (3.11), temos $A^c = \emptyset$ e logo
$$\mathbb{P}(\emptyset) = 1 - \mathbb{P}(\Omega) = 1 - 1 = 0, \tag{3.16}$$
e temos (3.12).

Nas condições de (3.13), temos $B = A \cup (B \setminus A)$, com A e $B \setminus A$ claramente disjuntos. Veja a Figura 3.1. De (3.10),
$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \tag{3.17}$$
e (3.13) segue.

Note que $A \cup B = A \cup (B \setminus A)$, onde a última união é disjunta (isto é, envolve subconjuntos disjuntos, quais sejam, A e $B \setminus A$). Veja a Figura 3.2. Aplicando então (3.10), temos

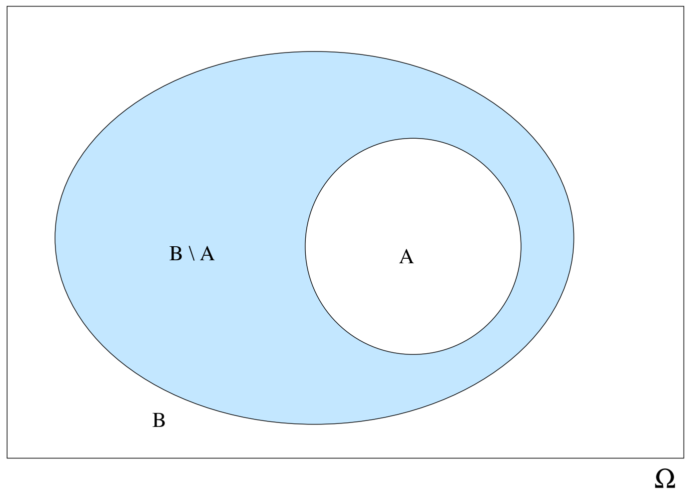

Figura 3.1 Retângulo representa o espaço amostral; elipse é B; círculo, A; região sombreada é $B \setminus A$.

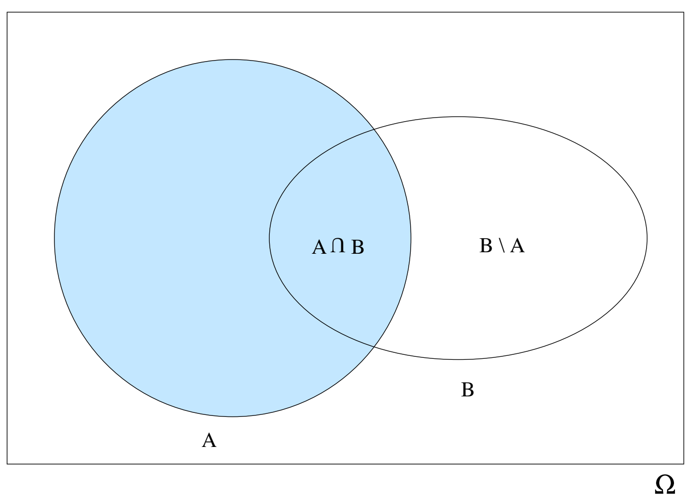

Figura 3.2 Círculo representa A; elipse, B; parte sombreada de B é $A \cap B$; parte não sombreada é $B \setminus A$.

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A). \tag{3.18}$$

Agora, note que $B \setminus A = B \setminus (A \cap B)$. Veja a Figura 3.2. Aplicando agora (3.13) com $A \cap B$ no lugar de A (note que $A \cap B \subset B$), temos

$$\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A \cap B), \tag{3.19}$$

e (3.14) segue de (3.18) e (3.19).

A aditividade da probabilidade (3.10) se estende para mais eventos. Dados n eventos disjuntos $A_1, A_2, \ldots, A_n \in \mathcal{E}$ (isto é, $A_i \cap A_j = \emptyset$ para todo $i, j = 1, \ldots, n$ com $i \neq j$), temos

$$\mathbb{P}\left(\cup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i). \tag{3.20}$$

Isto pode ser provado por indução em n, usando o caso $n = 2$ conhecido e o fato que o evento $\cup_{i=1}^{n-1} A_i$ e o evento A_n são disjuntos.

Voltando ao Exemplo 3.1, vemos que partindo da atribuição (3.1) (que, seguindo a idéia de que probabilidades devem ser atribuídas a eventos (subconjuntos) de Ω e não a seus pontos, devia ser denotada

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{3\}) = \mathbb{P}(\{4\}) = \mathbb{P}(\{5\}) = \mathbb{P}(\{6\}) = \frac{1}{6}), \tag{3.21}$$

e usando (3.20), temos que para todo $A \in \mathcal{E}$

$$\mathbb{P}(A) = \mathbb{P}\left(\cup_{i \in A} \{i\}\right) = \sum_{i \in A} \mathbb{P}(\{i\}) = \frac{1}{6} \sum_{i \in A} 1 = \frac{1}{6} \# A, \tag{3.22}$$

onde $\#A$ é, como já vimos antes, a cardinalidade ou número de elementos de A. Vemos que é suficiente neste caso atribuir probabilidades aos conjuntos unitários. As probabilidades dos demais eventos ficam determinadas pela aditividade.

Espaços amostrais finitos

O argumento que acabamos de usar pode ser usado em geral para o caso de espaços amostrais finitos. Seja Ω o espaço amostral do modelo probabilístico de um experimento aleatório. Suponha que Ω seja finito:

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}.$$

Podemos então tomar o conjunto das partes de Ω como a classe de eventos $\mathcal{E}$. A atribuição de probabilidades pode ser feita aos conjuntos unitários, da seguinte forma, que é geral.

Sejam $p_1, p_2, \dots, p_N$ números não negativos somando 1. Isto é,

$$p_i \geq 0 \quad \text{para } i = 1, 2, \dots, N \quad \text{e} \quad \sum_{i=1}^{N} p_i = 1. \tag{3.23}$$

Então, se fizermos a atribuição

$$\mathbb{P}(\{\omega_i\}) = p_i \quad \text{para } i = 1, 2, \dots, N, \tag{3.24}$$

então a probabilidade de um evento qualquer $A \in \mathcal{E}$ fica determinada pelas propriedades (3.12) e (3.10), e como em (3.22), obtemos para todo $A \in \mathcal{E}$ (ou, no caso, para todo $A \subset \Omega$, o que dá no mesmo)

$$\mathbb{P}(A) = \mathbb{P}\left(\cup_{i:\omega_i \in A}\{\omega_i\}\right) = \sum_{i:\,\omega_i \in A} \mathbb{P}(\{\omega_i\}) = \sum_{i:\,\omega_i \in A} p_i. \tag{3.25}$$

Note que as propriedades definidoras de uma probabilidade (funçao de $\mathcal{E}$ em $[0, 1]$ satisfazendo (3.9) e (3.10)) estão satisfeitas por (3.23) e (3.25)).

O Exemplo 3.1 é um caso particular em que $N = 6$ e $p_i \equiv 1/6$.

Espaços amostrais infinitos enumeráveis

Uma atribuição semelhante pode ser feita no caso em que Ω é infinito enumerável, isto é,

$$\Omega = \{\omega_1, \omega_2, \dots\}.$$

Neste caso, sendo $p_1, p_2, \dots$ uma seqüência infinita de números satisfazendo

$$p_i \geq 0 \quad \text{para todo } i = 1, 2, \dots \quad \text{e} \quad \sum_{i=1}^{\infty} p_i = 1, \tag{3.26}$$

podemos fazer a atribuição

$$\mathbb{P}(\{\omega_i\}) = p_i \quad \text{para todo } i = 1, 2, \dots, \tag{3.27}$$

e com isto para todo $A \in \mathcal{E}$ (que pode continuar sendo o conjunto das partes de Ω), temos $\mathbb{P}(A)$ exatamente como em (3.25) (mas note que neste caso precisamos – na segunda igualdade — de uma versão de (3.20) para infinitos eventos disjuntos, o que é uma propriedade adicional que impomos à probabilidade).

Exemplo 3.5 *Suponha que estejamos num jogo em que se lança uma moeda honesta até sair a primeira cara. Se o número de lançamentos necessários for par, ganhamos; se for ímpar, perdemos. Qual a probabilidade de vitória?*

Vamos construir um modelo probabilístico para o experimento aleatório que seria lançar uma moeda honesta até sair a primeira cara. O conjunto de possibilidades, ou espaço amostral, para o número de lançamentos seria

$$\Omega = \{1, 2, \ldots\},$$

todos os números naturais positivos, um conjunto infinito enumerável.

Vamos seguir a idéia acima e atribuir probabilidades aos subconjuntos unitários de Ω. Para fazê-lo, note que para que $\{i\}$ ocorra, é necessário e suficiente que os primeiros $i-1$ lançamentos resultem em coroa e o i-ésimo lançamento resulte em cara. Teremos então para $i = 1, 2, \ldots$

$$\mathbb{P}(\{i\}) = \left(\frac{1}{2}\right)^{i-1} \frac{1}{2} = \left(\frac{1}{2}\right)^{i}. \tag{3.28}$$

Note que o lado direito de (3.28) satisfaz (3.26) e logo, sendo vitória o evento $\{2, 4, 6, \ldots\}$, temos

$$\begin{aligned}\mathbb{P}(\textit{vitória}) = \mathbb{P}(\{2,4,6,\ldots\}) = \textstyle\sum_{i=2,4,6,\ldots} \left(\frac{1}{2}\right)^i = \sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^{2j} = \sum_{j=1}^{\infty} \left(\frac{1}{4}\right)^j \\ = \tfrac{1/4}{1-1/4} = \tfrac{1}{3}.\end{aligned}$$

Espaços amostrais não enumeráveis

Quando Ω é não enumerável, como no Exemplo 3.2, não basta em geral atribuir probabilidades aos eventos unitários. Normalmente, procuramos uma subclasse de $\mathcal{E}$ (que por sua vez, como já dissemos acima, neste caso não é em geral o conjunto das partes) e fazemos uma atribuição razoável aos eventos desta subclasse, e usamos as propriedades da probabilidade para obter as probabilidades dos demais eventos.

Nestas notas não veremos outros casos que não $\Omega =$ um subconjunto de $\mathbb{R}^n$, como um intervalo em $n = 1$, ou um hiperretângulo em $n \geq 2$. Vamos ver dois exemplos.

O primeiro exemplo seria o mesmo espaço amostral do Exemplo 3.2

$$\Omega = [0, 1].$$

Uma subclasse de eventos a que atribuir probabilidades inicialmente seriam os subintervalos de $[0,1]$. Como alternativa ou extensão do comprimento do subintervalo como a probabilidade do subintervalo, poderíamos também fazer a seguinte atribuição mais geral.

Vamos introduzir uma função contínua não decrescente $F:[0,1]\to[0,1]$ com a propriedade que $F(0)=0$ e $F(1)=1$. Então a atribuição

$$\mathbb{P}([a,b]) = F(b) - F(a)$$

para todo subintervalo $[a,b]\subset[0,1]$ define uma probabilidade neste caso (a subclasse é a classe dos subintervalos de $[0,1]$). Note que no Exemplo 3.2 $F(x)=x$, mas atribuições com $F(x)=x^n$, $n\geq 2$ também funcionam como atribuição de probabilidade (mas perderíamos a uniformidade), e mais genericamente qualquer F não decrescente com $F(0)=0$ e $F(1)=1$.

Suponha que $F(x)=\sqrt{x}$. Vamos calcular

$$\mathbb{P}([1/4,3/4]) = \sqrt{3/4}-\sqrt{1/4} = 0.37,$$

$$\begin{aligned}\mathbb{P}([1/8,1/3]\cup[2/5,4/7]) &= \mathbb{P}([1/8,1/3])+\mathbb{P}([2/5,4/7]) \\ &= (\sqrt{1/3}-\sqrt{1/8})+(\sqrt{4/7}-\sqrt{2/5}) = 0.35.\end{aligned}$$

Exemplo 3.6 *Suponha que nosso experimento aleatório seja escolher um ponto ao acaso de maneira uniforme do círculo unitário centrado na origem em $\mathbb{R}^2$, denotado $\mathcal{C}$.*

Neste caso $\Omega=\mathcal{C}=\{(x,y)\in\mathbb{R}^2 : x^2+y^2\leq 1\}$. A subclasse a ser considerada pode ser os retângulos $R=[a,b]\times[c,d]$ contidos em $\mathcal{C}$. Veja a Figura 3.3. Para estes, a atribuição natural (em vista da uniformidade) é

$$\mathbb{P}(R) = \frac{(b-a)(d-c)}{\pi}. \tag{3.29}$$

A partir desta atribuição, temos que para a maior parte dos subconjuntos C de $\mathcal{C}$ de interesse, temos

$$\mathbb{P}(C) = \frac{\text{área de } C}{\pi}, \tag{3.30}$$

estendendo (3.29).

Qual a probabilidade de o ponto escolhido pertencer ao quadrado inscrito em $\mathcal{C}$? Este é o retângulo $Q=[-\sqrt{2}/2,\sqrt{2}/2]\times[-\sqrt{2}/2,\sqrt{2}/2]$. Veja a Figura 3.4.

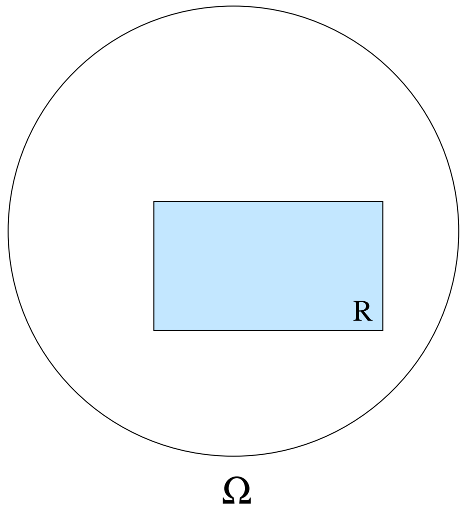

Figura 3.3 Região sombreada é retângulo $[a, b] \times [c, d]$.

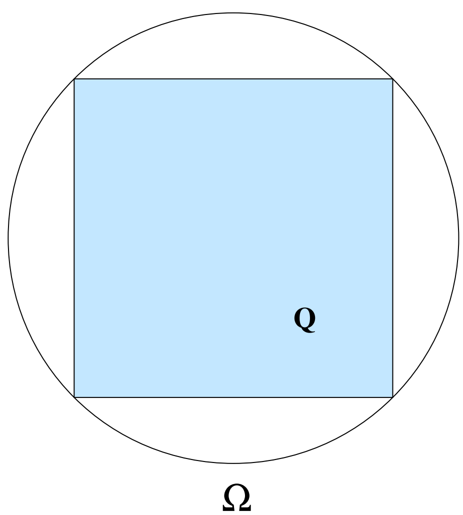

Figura 3.4 Região sombreada é o quadrado inscrito.

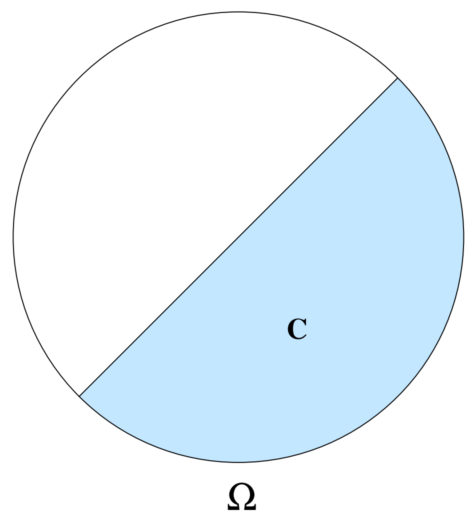

Figura 3.5 C é região sombreada.

Logo, de (3.29),

$$\mathbb{P}(Q) = \frac{2}{\pi}. \tag{3.31}$$

Qual a probabilidade de o ponto escolhido (X, Y) *ser tal que* $X > Y$*? Estamos querendo* $\mathbb{P}(C)$ *para* $C = \mathcal{C} \cap \{(x, y) \in \mathbb{R}^2 : x > y\}$*. Veja a Figura 3.5.*

De (3.30),

$$\mathbb{P}(C) = \frac{\pi/2}{\pi} = \frac{1}{2}. \tag{3.32}$$

Interpretação da probabilidade de um evento

Qual o significado da probabilidade de um evento? A seguinte é uma interpretação estatística, já apresentada no Exemplo 3.1. Dado um experimento aleatório qualquer e um seu evento A, a probabilidade de A seria o limite da freqüência de ocorrência de A em n repetições do experimento quando $n \to \infty$, se tal limite existir. Isto é, se $N_n(A)$ denotar o número de vezes em que A ocorre em n repetições do experimento, então

$$\mathbb{P}(A) = \lim_{n \to \infty} \frac{N_n(A)}{n}. \tag{3.33}$$

Isto pressupõe que o experimento seja (infinitamente) repetível e que o limite exista. Para esta e outras situações em que tais pressuposições não valham, há a interpretação subjetivista de que $\mathbb{P}(A)$ é o grau de crença que dado observador deposita a priori na ocorrência de A.

3.2 Espaços equiprováveis

Um caso particular importante dos modelos probabilísticos, de que o modelo para o dado equilibrado introduzido acima é um exemplo, é quando temos um espaço amostral finito

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$$

e cada evento unitário tem a mesma probabilidade:

$$\mathbb{P}(\{\omega_i\}) = \frac{1}{N} \text{ para todo } i = 1, 2, \dots, N. \tag{3.34}$$

Logo para qualquer evento A (que neste caso pode ser qualquer subconjunto de Ω):

$$\mathbb{P}(A) = \mathbb{P}\left(\cup_{\omega\in A}\{\omega\}\right) = \sum_{\omega\in A}\mathbb{P}(\{\omega\}) = \sum_{\omega\in A}\frac{1}{N} = \frac{1}{N}\sum_{\omega\in A}1 = \frac{\#A}{N} = \frac{\#A}{\#\Omega}, \tag{3.35}$$

onde $\#A$ é a cardinalidade de A, ou, em outras palavras, o número de elementos de A.

O cálculo de probabilidades nestes modelos se reduz pois essencialmente à contagem (do número de elementos dos eventos em questão).

No Exemplo 3.1, sejam os eventos

$$\begin{aligned} A &= \text{ o número lançado é par,} \\ B &= \text{ o número lançado é par ou maior do que 3,} \\ C &= \text{ o número lançado é par e maior do que 3.} \end{aligned}$$

Então

$$A = \{2, 4, 6\}, \quad B = \{2, 4, 5, 6\}, \quad C = \{4, 6\}.$$

Logo, de (3.35)

$$\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2}; \quad \mathbb{P}(B) = \frac{4}{6} = \frac{2}{3}; \quad \mathbb{P}(C) = \frac{2}{6} = \frac{1}{3}.$$

Exemplo 3.7 *Seja o experimento aleatório em que dois dados equilibrados são lançados um após o outro. Um modelo para este experimento seria um modelo equiprovável em que*

$$\begin{aligned}\Omega &= \{1,2,3,4,5,6\}^2 \\ &= \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\} \\ &= \{(1,1),(1,2),\ldots,(1,6), \\ &\quad (2,1),(2,2),\ldots,(2,6), \\ &\quad \ldots, \\ &\quad (6,1),(6,2),\ldots,(6,6)\}. \end{aligned} \tag{3.36}$$

Neste caso, $N = 36$, e $(i,j) \in \Omega$ indica os números do primeiro e segundo dados respectivamente.

Qual a probabilidade de que a soma dos números lançados seja 7?

O evento $A =$ a soma dos números lançados é 7 pode ser descrito como

$$A = \{(i,j) \in \Omega : i+j = 7\} = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}. \tag{3.37}$$

Logo, de (3.35)

$$\mathbb{P}(A) = \frac{6}{36} = \frac{1}{6}.$$

Uma forma de obter a cardinalidade de A é escrever Ω em forma de matriz

$$\Omega = (ij)_{\substack{i=1,\ldots,6 \\ j=1,\ldots,6}}$$

(um pouco como a partir da última igualdade em (3.36)), e notar que os elementos de A dispõem-se na diagonal secundária da matriz. Logo, A tem 6 elementos, que é o número de elementos da diagonal.

Amostragem aleatória em populações

Discutimos em seguida uma situação importante em estatística. Suponha que

$$\Pi = \{I_1, I_2 \ldots, I_M\}$$

seja uma população com M indivíduos (em que o j-ésimo indivíduo é indicado por I_j, $j = 1, 2, \ldots, M$).

Uma amostra aleatória de Π é grosso modo um subconjunto de Π escolhido aleatoriamente. O exemplo mais simples é a amostra aleatória (ou casual) simples de tamanho 1. Esta é formada pelo sorteio de 1 indivíduo de Π em que cada indivíduo tem a mesma chance de ser sorteado que os demais. Isto nos leva a considerar um modelo equiprovável para o sorteio em que $\Omega = \Pi$.

Exemplo 3.8 *Suponha que dada população tenha* 55 *mulheres e* 45 *homens. Se tomarmos uma amostra casual simples de tamanho* 1 *desta população, qual a probabilidade de sortearmos uma mulher?*

Está claro que o evento A = "indivíduo sorteado é uma mulher" tem cardinalidade 55*. Logo,*

$$\mathbb{P}(A) = \frac{55}{100} = 0.55.$$

Para amostras de tamanho maior do que 1, temos dois casos: amostragem com e sem reposição.

Amostra casual simples com reposição

Suponha que queiramos uma amostra de tamanho $n \geq 2$ escolhida da seguinte forma: sorteamos o primeiro indivíduo de Π para a amostra como na no caso $n = 1$; devolvemos o indivíduo sorteado à população, e repetimos o procedimento, e assim até o n-ésimo sorteio.

Um modelo para esta amostragem é um modelo equiprovável com

$$\begin{aligned} \Omega &= \Pi^n = \Pi \times \ldots \times \Pi \quad (n \text{ vezes}) \\ &= \{(I_{j_1}, \ldots, I_{j_n}) : (j_1, \ldots, j_n) \in \{1, \ldots, M\}^n\}. \end{aligned} \tag{3.38}$$

Note que pode haver repetições, ou seja, um mesmo indivíduo pode ser sorteado mais do que uma vez. Note ainda que

$$N = \#\Omega = M^n. \tag{3.39}$$

Exemplo 3.9 *Suponha que tomemos uma amostra casual simples de tamanho* 5 *com reposição da população do Exemplo 3.8. Neste caso, $N = \#\Omega = 100^5 = 10^{10}$. Sejam os eventos*

$$\begin{aligned} A &= \textit{não há nenhuma mulher na amostra}, \\ B &= \textit{há exatamente } 3 \textit{ mulheres na amostra}, \\ C &= \textit{as mulheres estão em maioria na amostra}. \end{aligned}$$

Para achar as probabilidades destes eventos, vamos determinar as cardinalidades de cada um deles.

Suponha que $\vec{\Pi}$ *e* Π_+ *sejam os subconjuntos de homens e mulheres de* Π *respectivamente. Temos que* $\#\vec{\Pi} = 45$ *e* $\#\Pi_+ = 55$*. Temos então que* A *é o subconjunto de* Ω *com amostras de apenas homens. Este subconjunto pode ser descrito como* $\vec{\Pi}^5$*, e logo*

$$\#A = \#\vec{\Pi}^5 = 45^5.$$

Portanto,

$$\mathbb{P}(A) = \frac{45^5}{100^5} = \left(\frac{45}{100}\right)^5 = 0.45^5 = 0.018.$$

Para achar a cardinalidade de B*, vamos considerar as posições ordenadas em que aparecem as 3 mulheres. Há* $\binom{5}{3}$ *possibilidades para as posições. Fixadas as posições das mulheres (o que fixa as posições dos 2 homens), a subamostra de mulheres que vai ocupar as 3 posições fixadas é qualquer uma de* Π_+^3*, e a subamostra de homens é qualquer uma de* $\vec{\Pi}^2$*. Logo*

$$\#B = \binom{5}{3} \times \#\vec{\Pi}^2 \times \#\Pi_+^3 = 10 \times 45^2 \times 55^3,$$

e portanto

$$\mathbb{P}(B) = \frac{10 \times 45^2 \times 55^3}{100^5} = 10\left(\frac{45}{100}\right)^2\left(\frac{55}{100}\right)^3 = 10 \times 0.45^2 \times 0.55^3 = 0.337.$$

Finalmente, $C = B \cup B' \cup B''$*, união disjunta em que*

$$\begin{aligned} B' &= \text{há exatamente 4 mulheres na amostra}, \\ B'' &= \text{há exatamente 5 mulheres na amostra}. \end{aligned}$$

Pela aditividade da probabilidade (3.20), basta acharmos $\mathbb{P}(B')$ *e* $\mathbb{P}(B'')$*. Estas são obtidas de forma similar a* $\mathbb{P}(B)$*.*

$$\begin{aligned} \mathbb{P}(B') &= \binom{5}{4}\left(\frac{45}{100}\right)^1\left(\frac{55}{100}\right)^4 = 5 \times 0.45 \times 0.55^4 = 0.206, \\ \mathbb{P}(B'') &= \left(\frac{55}{100}\right)^5 = 0.55^5 = 0.050 \end{aligned}$$

(verifique), e, da aditividade,

$$\mathbb{P}(C) = \mathbb{P}(B) + \mathbb{P}(B') + \mathbb{P}(B'') = 0.593.$$

Amostra casual simples sem reposição

Esta amostragem é semelhante à com reposição, com uma diferença importante: não há a devolução do indivíduo sorteado à população no final de cada sorteio. Desta forma, temos um modelo equiprovável com

$$\Omega = \left\{ (I_{j_1}, \ldots, I_{j_n}) : (j_1, \ldots, j_n) \in \{1, \ldots, M\}^n ;\, j_i \neq j_k \text{ se } i \neq k \right\},$$

de forma que não há repetições: cada indivíduo aperece no máximo uma vez na amostra. Isto naturalmente obriga a que $n \leq M$. Vamos também estipular que amostras dos mesmos indivíduos em ordem distinta só contam uma vez; desta forma, a cardinalidade de Ω é o número de escolhas de um grupo de n indivíduos distintos numa população de M indivíduos. Logo

$$N = \#\Omega = \binom{M}{n}, \tag{3.40}$$

a *combinação* de M, n a n.

Exemplo 3.10 *Vamos tomar a mesma população, tamanho de amostra e eventos do Exemplo 3.9, mas com amostragem sem reposição.*

Para determinar $\#A$, note que o número de amostras com apenas homens é o número de escolhas de 5 *indivíduos distintos de uma população, $\vec{\Pi}$, com* 45 *indivíduos. Logo,*

$$\#A = \binom{45}{5}$$

e portanto

$$\mathbb{P}(A) = \frac{\binom{45}{5}}{\binom{100}{5}} = \frac{\frac{45!}{40!\,5!}}{\frac{100!}{95!\,5!}} = \frac{45 \times 44 \times 43 \times 42 \times 41}{100 \times 99 \times 98 \times 97 \times 96} = 0.016.$$

Em B temos uma escolha de 3 *indivíduos de Π_+ e uma escolha de* 2 *indivíduos de $\vec{\Pi}$. O número de possibilidades da primeira é $\binom{55}{3}$; da segunda, $\binom{45}{2}$. No total*

$$\#B = \binom{45}{2}\binom{55}{3}$$

e portanto

$$\mathbb{P}(B) = \frac{\binom{45}{2}\binom{55}{3}}{\binom{100}{5}} = \frac{\frac{45!}{40!\,5!}}{\frac{100!}{95!\,5!}} = 0.345.$$

Para calcular $\mathbb{P}(C)$, proceda de forma similar ao que fizemos antes, notando que $\#B' = \binom{45}{1}\binom{55}{4}$ e $\#B'' = \binom{55}{5}$.

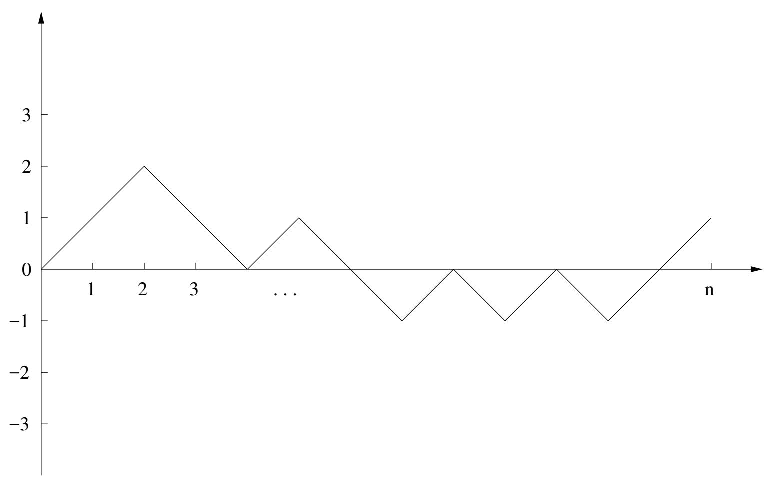

Figura 3.6 Trajetória típica dos n passos do caminhante; abscissa representa número de passos; ordenada, a posição.

3.2.1 Passeio aleatório

Uma pessoa sai caminhando de um bar, a princípio em direção a sua casa, a alguns quarteirões, na mesma rua. Mas suponha que o seu estado seja tal que cada passo tenha a mesma probabilidade de ser na direção correta ou na direção oposta. Onde se encontrará esta pessoa após n passos? $(n \geq 1)$

Vamos descrever um modelo probabilístico equiprovável para esta situação. Vamos começar supondo que cada passo tem sempre o mesmo comprimento 1. O espaço amostral $\Omega = \Omega_n$ consistirá de todas as possíveis trajetórias de n passos de tamanho 1, cada passo sendo para baixo ou para cima. Veja a Figura 3.6.

Para determinar a cardinalidade de Ω_n, note que cada trajeória é determinada pela seqüência de n passos, sendo que cada um tem duas possibilidades (para baixo ou para cima). Temos então um total de 2^n tais seqüências, e portanto, 2^n trajetórias em Ω_n, cada qual com probabilidade 2^{-n}.

Vamos denotar uma trajetória de Ω_n por $S = S_n$, e por $S(1)$, $S(2)$, …, $S(n)$ as sucessivas posições visitadas por S.

Queremos pois responder à pergunta: qual é a probabilidade de que $S(n) = k$? Note que as possibilidades para k vão desde de $-n$ (todos os passos para baixo), até n (todos os passos para cima), com todas as possibilidades intermediárias de mesma paridade que n (isto é, todo k intermediário tal que $k + n$ seja par).

Para calcular a cardinalidade do evento $\{S(n) = k\}$, onde k satisfaz as restrições acima, vamos considerar

$$L = \text{o número de passos para cima da trajetória } S.$$

De fato, $S(n)$ pode ser obtido da diferença entre o número de passos para cima, L, e o número de passos para baixo, $n - L$. Temos que

$$S(n) = k \Leftrightarrow L - (n - L) = 2L - n = k \Leftrightarrow L = \frac{n+k}{2}.$$

Logo $\{S(n) = k\} = \{L = (n + k)/2\}$, logo $\#\{S(n) = k\}$ é o quantidade de escolhas de $(n + k)/2$ passos, a serem dados para cima, de um total de n passos. Isto é dado pela combinação de n, $(n + k)/2$ a $(n + k)/2$:

$$\binom{n}{\frac{n+k}{2}},$$

e logo

$$\mathbb{P}(S(n) = k) = \frac{\#\{S(n) = k\}}{\#\Omega_n} = \binom{n}{\frac{n+k}{2}} 2^{-n},$$

$k = -n, -n + 2, \ldots, n - 2, n$.

Volta à origem

Se a pessoa fizer a caminhada descrita acima, que chamamos de *passeio aleatório*, indefinidamente, será que, mais cedo ou mais tarde, ela acabará voltando à origem (isto é, ao bar)? Pela restrição de paridade que já discutimos, se isto ocorrer, deverá ser num instante par (digamos que um passo é dado a cada instante), e qualquer tal instante é uma possibilidade. Seja

$$T = \min\{n > 0 : S(n) = 0\}.$$

Vamos achar $\mathbb{P}(T = 2k)$ para um valor arbitrário de $k = 1, 2, \ldots$

O evento $\{T = 2k\}$ será tratado como evento do espaço de probabilidades equiprovável de trajetórias de comprimento $2k$ passos, e pode ser dividido em dois eventos simétricos: aquele no qual o primeiro passo é para cima — e logo, necessariamente o último passo é para baixo —; e aquele no qual o primeiro passo é para baixo — e logo, necessariamente o último passo é para cima. Denotemos o primeiro evento por Γ_+, e o segundo por Γ_-. Pela simetria, é claro que

$$\#\{T = 2k\} = \#\Gamma_+ + \#\Gamma_- = 2\#\Gamma_+. \tag{3.41}$$

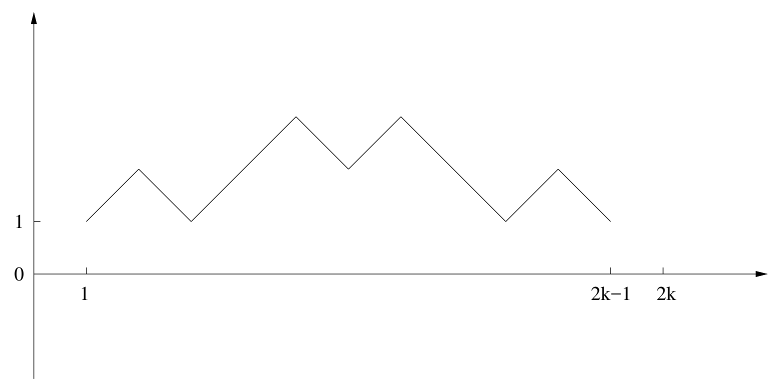

Figura 3.7 Entre os instantes 1 e $2k-1$, trajetórias de Γ_+ não tocam a abscissa.

Podemos descrever Γ_+ como

$$\begin{aligned}\Gamma_+ &= \{\text{trajetórias } S:\ S(1) = S(2k-1) = 1 \\ &\qquad \text{e } S(j) > 0 \text{ para todo } 1 < j < 2k-1\}.\end{aligned}$$

Veja a Figura 3.7.

Seja agora o evento

$$\tilde{\Gamma} = \{\text{trajetórias } S:\ S(1) = S(2k-1) = 1\}.$$

Então podemos calcular a cardinalidade de $\tilde{\Gamma}$ como a de $\{S(n) = k\}$ acima (de fato, é visível que $\#\tilde{\Gamma} = \#\{S(2k-2) = 0\}$), obtendo

$$\#\tilde{\Gamma} = \binom{2k-2}{k-1}. \tag{3.42}$$

Observamos agora que $\tilde{\Gamma}$ se decompõe de forma disjunta em Γ_+ e

$$\begin{aligned}\Gamma_0 &= \{\text{trajetórias } S:\ S(1) = S(2k-1) = 1, \\ &\qquad \text{e } S(j) = 0 \text{ para algum } 1 < j < 2k-1\},\end{aligned}$$

isto é, Γ_0 consiste das trajetórias de $\tilde{\Gamma}$ que tocam ou cruzam a abscissa entre os instantes 1 e $2k-1$.

Vamos agora apresentar um argumento que nos dá $\#\Gamma_0$. Ele usa o *princípio da reflexão*.

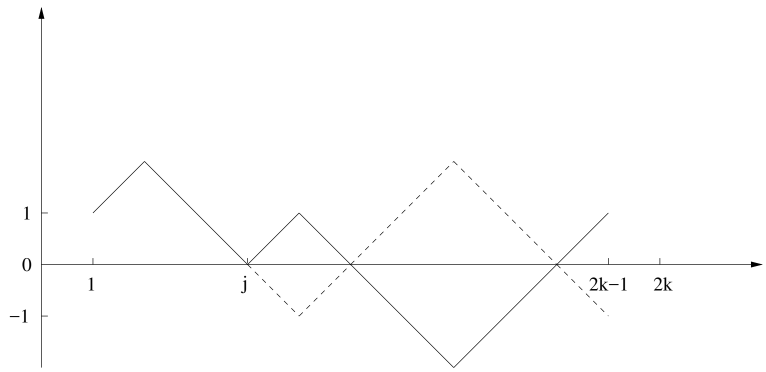

Figura 3.8 S é a trajeória cheia; S' coincide com S até j, e segue a porção tracejada a partir daí. A porção tracejada é o reflexo na abscissa da porção de S a partir de j.

Dada uma trajetória S de Γ_0, seja j o primeiro instante após 1 em que a abscissa é tocada, e considere a trajetória S' que coincide com S até o instante j e, a partir de j, segue o *reflexo* na abscissa da porção de S a partir de j. Veja a Figura 3.8.

Note agora que S' é uma trajetória que conecta $(1, 1)$ a $(2k - 1, -1)$. Reciprocamente, toda trajetória conectando $(1, 1)$ a $(2k - 1, -1)$ pode ser obtida desta forma (pois tal trajetória tem de cruzar a abscissa; seja j o instante do primeiro cruzamento; faça a reflexão e composição como antes para obter uma trajetória conectando $(1, 1)$ a $(2k-1, 1)$ e tocando a abscissa entre 2 e $2k - 2$).

A conclusão é o *princípio da reflexão*:

$$\#\Gamma_0 = \#\Gamma',$$

onde

$$\Gamma' = \{\text{trajetórias } S' \text{ conectando } (1, 1) \text{ a } (2k - 1, -1)\}.$$

Note que não há nenhuma restrição no meio das trajetórias de Γ'. Para obter $\#\Gamma'$, observamos que a única restrição em Γ' é que dos $2k - 2$ passos (entre os instantes 1 e $2k - 1$), exatamente k devem ser para baixo (e $k - 2$ para cima). Logo

$$\#\Gamma_0 = \#\Gamma' = \binom{2k-2}{k}, \tag{3.43}$$

e de (3.41), (3.42) e (3.43)

$$\#\{T = 2k\} = 2\left[\#\tilde{\Gamma} - \#\Gamma_0\right] = 2\left[\binom{2k-2}{k-1} - \binom{2k-2}{k}\right] = \frac{1}{2k-1}\binom{2k}{k}. \tag{3.44}$$

Portanto

$$\mathbb{P}(T = 2k) = \frac{\#\{T = 2k\}}{\#\Omega_{2k}} = \frac{1}{2k-1}\binom{2k}{k}2^{-2k},\ k \geq 1. \tag{3.45}$$

Agora, se somarmos a expressão em (3.45) para $k = 1, 2, \ldots$ teremos a probabilidade de que o caminhante volte à origem eventualmente. (Neste raciocínio, os eventos $\{T = 2k\}$, $k \geq 1$, devem ser vistos como eventos – disjuntos – do espaço amostral de todas as trajetórias infinitas.)

Veremos no Apêndice B a estas notas, por um método que de fato prescinde do princípio da reflexão (mas recorre a uma outra propriedade importante do passeio aleatório, a saber, a perda de memória, ou propriedade de Markov), que

$$\mathbb{P}(\text{retorno eventual à origem}) = \sum_{k=1}^{\infty} \mathbb{P}(T = 2k) = 1, \tag{3.46}$$

e temos que o retorno se dá com probabilidade 1.

Observação 3.11 *Apesar da apresentação pitoresca feita nesta subseção, o passeio aleatório serve de modelo para muitas situações práticas importantes. Ele é um modelo microscópico para o movimento de partículas em fenômenos físicos, como o movimento Browniano. Ele entra na modelagem de preços de ativos financeiros em mercados equilibrados. Há muitos outros exemplos em diversas áreas.*

3.3 Outros exemplos

Exemplo 3.12 (Aniversários) *Numa classe com n alunos, qual a probabilidade de pelo menos dois deles fazerem aniversário no mesmo dia?*

Para responder a esta pergunta, vamos considerar o evento complementar

$$A = \{\textit{ninguém faz aniversário no mesmo dia}\}.$$

Para achar $\mathbb{P}(A)$, vamos supor que os aniversários da classe são uma amostra casual simples de tamanho n com reposição de

$$\Pi = \{1, 2, \ldots, 365\},$$

os diferentes dias do ano enumerados de alguma forma (ignorando 29 *de Fevereiro).*

Desta forma, como acima $\Omega = \Pi^n$, e logo $\#\Omega = 365^n$ (veja (3.39)).

Para obter a cardinalidade de A, note que o primeiro aniversário a ser sorteado tem 365 *possibilidades; o segundo,* 364 *(já que não pode coincidir com o primeiro para estar em A); o terceiro,* 363*, e assim por diante, até o n-ésimo, que tem $365 - n + 1$ possibilidades (vamos supor que $n \leq 365$).*

Então,

$$\#A = 365 \times 364 \times 363 \times \ldots \times (365 - n + 1) =: (365)_n,$$

o arranjo *de* 365 *n a n. Logo,*

$$\mathbb{P}(A) = \frac{(365)_n}{365^n} = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\ldots\left(1 - \frac{n-1}{365}\right).$$

Verifique que para $n \geq 23$, $\mathbb{P}(A) < 1/2$, e logo podemos concluir que numa classe de pelo menos 23 *alunos, a probabilidade de haver pelo menos uma coincidência de aniversário é de pelo menos* 50%.

Exemplo 3.13 (Peixes) *Num lago há um número N de peixes de certa espécie. Uma equipe faz uma pescaria de K peixes da espécie em questão. Estes são marcados, e devolvidos ao lago. Pouco tempo depois uma nova pescaria de $n \leq K$ peixes é feita e descobre-se que k destes estão marcados. O que os números K, n e k nos dizem sobre N?*

Vamos supor que a segunda pescaria nos dá uma amostra casual simples sem reposição de tamanho n de uma população de N peixes em que K estão marcados e $N - K$ não têm marca. (Estamos supondo que não houve mudanças na população entre as duas pescarias.)

Seja X o número de peixes marcados na amostra. Vamos calcular $\mathbb{P}(X = k)$, e para isto precisamos achar $\#\{X = k\}$.

Em $\{X = k\}$, de K peixes marcados, escolhemos k, num total de $\binom{K}{k}$ possibilidades, e de $N - K$ sem marca, escolhemos $n - k$, num total de $\binom{N-K}{n-k}$ possibilidades. Logo

$$\#\{X = k\} = \binom{K}{k}\binom{N-K}{n-k}$$

e portanto

$$\mathbb{P}(X=k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} =: q_N,$$

onde o denominador é $\#\Omega$ *(veja (3.40)).*

Tomaremos como estimador de N*, que denotaremos* $\hat{N}$*, o valor que maximizar* q_N*, com* K, n *e* k *fixos. Para isto vamos tomar o quociente*

$$\frac{q_{N+1}}{q_N} = \frac{\binom{N+1-K}{n-k}}{\binom{N+1}{n}} \Bigg/ \frac{\binom{N-K}{n-k}}{\binom{N}{n}} = \frac{1-\frac{K}{N+1}}{1-\frac{K-k}{N+1-n}}.$$

Notando que $N \geq K+n-k$ *e* $n \geq k$*, temos que o quociente acima é* ≥ 1 *se e só se* $N \leq \frac{n}{k}K - 1$*. Concluimos que* N *que maximiza* $\mathbb{P}(X=k)$ *pode ser tomado como o maior dentre* $\left\lfloor \frac{n}{k}K \right\rfloor$ *e* $K+n-k$*:*

$$\hat{N} = \left\lfloor \frac{n}{k}K \right\rfloor \vee (K+n-k), \tag{3.47}$$

onde $\lfloor \cdot \rfloor$ *indica a parte inteira, e* $\vee$ *o máximo.*

Suponha que $K = n = 1000$ *e* $k = 100$*. Substituindo em (3.47):*

$$\hat{N} = 10000 \tag{3.48}$$

seria a nossa estimativa para N*.*

Exemplo 3.14 (Chaves) *Uma pessoa tem um molho com* n *chaves das quais só uma abre sua porta. Ao chegar em casa, ela vai testando as chaves ao acaso (sem reposição), até achar a chave correta e abrir a porta. Qual a probabilidade de ela ser bem sucedida na* k*-ésima tentativa? (*$k = 1, \ldots, n$*)*

Vamos modelar esta situação por um espaço equiprovável em que os resultados são todas as possíveis ordenações das n *chaves. Temos então que* $\#\Omega = n!$*. No evento em questão, digamos* A*, a chave que abre a porta deve aparecer na* k*-ésima posição, e as demais* $n-1$ *chaves aparecem em qualquer ordem. Logo* $\#A = (n-1)!$*, e*

$$\mathbb{P}(A) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

(Este argumento parece supor que a pessoa continua a tentar as chaves mesmo depois de abrir a porta, mas isto não é relevante para a solução.)

Exemplo 3.15 (Máquinas fotográficas) *Um repórter fotográfico leva 2 máquinas fotográficas a tiracolo para a cobertura de um acontecimento. Cada máquina tem capacidade para tirar n fotos. Toda vez que o repórter quer tirar uma foto, ele pega uma das 2 máquinas ao acaso e tenta tirar uma foto, e repete este procedimento indefinidamente. Quando ele notar pela primeira vez que se esgotou a capacidade de uma máquina, qual a probabilidade de que a outra máquina esteja com capacidade k? ($k = 0, 1, \ldots, n$)*

O evento em questão ocorrerá com a máquina da direita se ela for a selecionada na escolha $2n-k+1$, e nas $2n-k$ escolhas anteriores ela aparecer n vezes (em qualquer ordem), e a máquina da esquerda aparecer $n-k$ vezes. A probabilidade disto é pois

$$\binom{2n-k}{n} 2^{-(2n-k+1)} \tag{3.49}$$

(o quociente em que o denominador é o número total de possibilidades nas primeiras $2n-k+1$ escolhas, e numerador é o número de possibilidades em que o evento em questão ocorrerá com a máquina da direita).

Como o evento em questão pode ocorrer também com a máquina da esquerda, e por simetria, a probabilidade disto é também (3.49), temos que a probabilidade desejada é

$$2\binom{2n-k}{k} 2^{-(2n-k+1)} = \binom{2n-k}{n} 2^{-(2n-k)}, \; k = 0, 1, \ldots, n \tag{3.50}$$

3.4 Condicionamento e independência

Suponha que na observação de um fenômeno aleatório, tenhamos informação parcial sobre o resultado, isto é, saibamos que dado evento ocorreu. Como isto afeta as chances relativas da ocorrência de um outro evento?

Seja $\Omega, \mathcal{E}, \mathbb{P}$ o espaço amostral, classe de eventos, e a probabilidade descrevendo o experimento original, respectivamente. Uma forma de vermos a nova situação, após sabermos que um evento $A \in \mathcal{E}$ ocorreu, é substituirmos Ω por $\Omega_A = A$, $\mathcal{E}_A = \{A \cap B : \, B \in \mathcal{E}\}$, e $\mathbb{P}$ por $\mathbb{P}_A = \mathbb{P}/\mathbb{P}(A)$, de forma que para $A' \in \mathcal{E}_A$, com $A' = A \cap B$ para algum $B \in \mathcal{E}$, teremos

$$\mathbb{P}_A(A') = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}. \tag{3.51}$$

(Verifique que $\mathbb{P}_A$ é de fato uma probabilidade em $(\Omega_A, \mathcal{E}_A)$, isto é, $\mathbb{P}_A : \mathcal{E}_A \to [0,1]$ é tal que $\mathbb{P}_A(\Omega_A) = 1$ e, para $A', B' \in \mathcal{E}_A$ disjuntos, temos $\mathbb{P}_A(A' \cup B') = \mathbb{P}_A(A') + \mathbb{P}_A(B')$.)

Uma forma mais conveniente, sem necessidade de passarmos a outro espaço amostral e espaço de eventos, é definir em $(\Omega, \mathcal{E})$ a *probabilidade condicionada em* A (ou *probabilidade condicional dado* A): para todo $B \in \mathcal{E}$,

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}. \tag{3.52}$$

Note que é a mesma expressão em (3.51).

Observação 3.16 *O lado direito de (3.52) (e (3.51)) só faz sentido a priori se* $\mathbb{P}(A) > 0$. *Quando* $\mathbb{P}(A) = 0$, *podemos definir* $\mathbb{P}(\cdot|A)$ *de forma arbitrária. Uma escolha conveniente neste caso é* $\mathbb{P}(\cdot|A) = \mathbb{P}(\cdot)$, *isto é,* $\mathbb{P}(B|A) = \mathbb{P}(B)$ *para todo* $B \in \mathcal{E}$, *se* $\mathbb{P}(A) = 0$.

Observação 3.17 *Verifique que* $\mathbb{P}(\cdot|A)$ *é uma probabilidade em* $(\Omega, \mathcal{E})$ *para todo* $A \in \mathcal{E}$.

Contudo, fixado $B \in \mathcal{E}$, $\mathbb{P}(B|\cdot)$ não *é em geral uma probabilidade em* $(\Omega, \mathcal{E})$.

No Exemplo 3.1 do lançamento de um dado, sejam os eventos

$$\begin{aligned} A &= \{\text{o número lançado é par}\}, & (3.53)\\ B &= \{\text{o número lançado é maior do que } 3\}. & (3.54) \end{aligned}$$

Então

$$\begin{aligned} \mathbb{P}(B|A) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(\{4,6\})}{\mathbb{P}(\{2,4,6\})} = \frac{2/6}{3/6} = \frac{2}{3}, & (3.55)\\ \mathbb{P}(B|A^c) &= \frac{\mathbb{P}(A^c \cap B)}{\mathbb{P}(A^c)} = \frac{\mathbb{P}(\{5\})}{\mathbb{P}(\{1,3,5\})} = \frac{1/6}{3/6} = \frac{1}{3} & (3.56) \end{aligned}$$

Da definição de probabilidade condicional segue-se imediatamente a chamada *regra do produto*: dados $A, B \in \mathcal{E}$

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\,\mathbb{P}(B) = \mathbb{P}(A)\,\mathbb{P}(B|A). \tag{3.57}$$

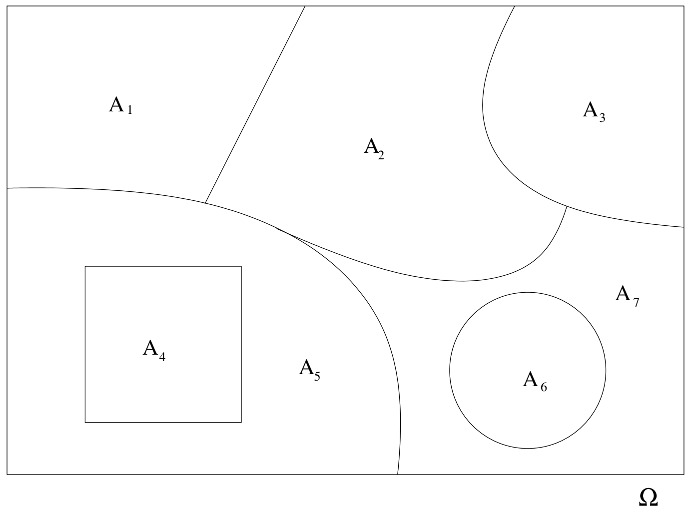

Figura 3.9 Partição do espaço amostral ($n = 7$).

Regra da probabilidade total

Uma aplicação importante de (3.57) é em estabelecer a *regra da probabilidade total*. Para isto comecemos por definir uma *partição* do espaço amostral. Sejam $A_1, A_2, \ldots, A_n \in \mathcal{E}$ eventos disjuntos e *exaustivos*, isto é, além de $A_i \cap A_j = \emptyset$ para todo $i, j = 1, 2, \ldots, n$ com $i \neq j$, temos

$$\cup_{i=1}^{n} A_i = \Omega.$$

Dizemos que $\{A_1, A_2, \ldots, A_n\}$ são uma partição (de Ω). Veja a Figura 3.9.

Neste caso, para qualquer $B \in \mathcal{E}$, temos

$$B = \cup_{i=1}^{n} \{A_i \cap B\},$$

uma união claramente disjunta. Veja a Figura 3.10.

Da aditividade da probabilidade (3.20), temos

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i \cap B), \tag{3.58}$$

e da regra do produto, obtemos a regra da probabilidade total:

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B|A_i)\mathbb{P}(A_i). \tag{3.59}$$

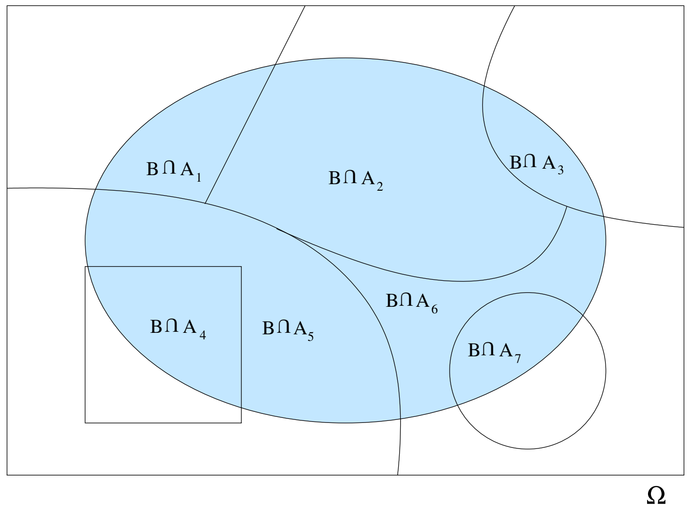

Figura 3.10 Elipse sombreada representa B.

No Exemplo 3.1, sejam A e B os eventos dados em (3.53) e (3.54), respectivamente. Temos então que $\{A, A^c\}$ é uma partição. Logo, da regra da probabilidade total,

$$\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c) = \frac{2}{3}\frac{1}{2} + \frac{1}{3}\frac{1}{2} = \frac{1}{2}. \qquad (3.60)$$

Neste caso, o cálculo direto de $\mathbb{P}(B)$ é mais natural e simples, mas em certas situações, é natural ou conveniente definir uma partição e usar a regra da probabilidade total. Isto normalmente é o caso quando o experimento aleatório em questão consiste de estágios, os resultados de um estágio servindo como partição para avaliar eventos do estágio seguinte.

Exemplo 3.18 *Um exemplo típico é a amostragem casual simples sem reposição. Suponha que numa urna haja K bolas azuis e M bolas brancas, e que lhe retiramos 2 bolas sem reposição. Este experimento tem dois estágios, que são as duas retiradas. Sejam os eventos*

$$\begin{aligned} A_1 &= \{\textit{sai bola azul na 1a. retirada}\}, \\ A_2 &= \{\textit{sai bola azul na 2a. retirada}\}. \end{aligned}$$

Ache $\mathbb{P}(A_2)$.

O cálculo de $\mathbb{P}(A_1)$ é claro: de $K + M$ possibilidades, tem que ocorrer uma de K possibilidades favoráveis a A_1. Então,

$$\mathbb{P}(A_1) = \frac{K}{K+M}. \tag{3.61}$$

O cálculo de $\mathbb{P}(A_2)$ é mais complicado, porque o que ocorre na 2a. retirada depende do que ocorre na 1a. retirada. A regra da probabilidade total é uma chave para a solução:

$$\mathbb{P}(A_2) = \mathbb{P}(A_2|A_1)\mathbb{P}(A_1) + \mathbb{P}(A_2|A_1^c)\mathbb{P}(A_1^c). \tag{3.62}$$

Note que $\{A_1, A_1^c\}$ é uma partição, e que

$$\mathbb{P}(A_2|A_1) = \frac{K-1}{K+M-1}, \tag{3.63}$$

pois, dado A_1, para a 2a. retirada temos $K-1$ bolas azuis e M bolas brancas na urna, logo $K-1$ possibilidades favoráveis a A_2 de $M+K-1$ possibilidades no total. Da mesma forma,

$$\mathbb{P}(A_2|A_1^c) = \frac{K}{K+M-1}, \tag{3.64}$$

e substituindo (3.63) e (3.64) em (3.62),

$$\begin{aligned}
\mathbb{P}(A_2) &= \frac{K-1}{K+M-1}\frac{K}{K+M} + \frac{K}{K+M-1}\frac{M}{K+M} \\
&= \frac{(K-1)K+KM}{(K+M)(K+M-1)} = \frac{K(K+M-1)}{(K+M)(K+M-1)} \\
&= \frac{K}{K+M}.
\end{aligned}$$

Uma forma simples de representar os eventos e probabilidades envolvidas na regra da probabilidade total é através de um diagrama chamado *árvore de probabilidades*.

A árvore de probabilidades é um diagrama em forma de árvore, com uma raiz, a partir da qual partem ramos indicando os eventos (de uma partição) do 1o. estágio; a partir de cada um destes, saem ramos indicando eventos do 2o. estágio, e assim por diante, até representar eventos do último estágio. Sobre cada ramo, indicamos a probabilidade condicional do evento indicado dada a seqüência de eventos ocorridos até ali.

Para calcular a probabilidade de determinado evento de certo estágio, localizamos todas as posições deste evento na árvore naquele estágio. Para cada posição, multiplicamos as probabilidades que encontramos sobre ramos no caminho desde a raiz até a posição em questão. Finalmente, somamos os produtos sobre todas as posições do evento em questão.

Na Figura 3.11 descrevemos de forma genérica um experimento em 3 estágios, em que cada estágio tem 3 eventos (em geral, o número de eventos pode variar a cada estágio).

Na Figura 3.12, representamos a situação do Exemplo 3.18 com $K = 6$ e $M = 4$. Vamos calcular $\mathbb{P}(A_2)$ usando esta árvore.

Temos A_2 em duas posições no segundo estágio (segunda retirada), indicadas por círculos pontilhados. Multiplicando as probabilidades que encontramos nos ramos do caminho da raiz até a primeira posição, temos

$$0.6 \times 0.56 = 0.33.$$

Fazendo o mesmo no caminho da raiz até a segunda posição, temos

$$0.4 \times 0.67 = 0.27.$$

Somando, vem

$$\mathbb{P}(A_2) = 0.6. \tag{3.65}$$

Exemplo 3.19 *Um meteorologista prevê corretamente o tempo de certa localidade em* 80% *dos dias de sol, e em* 60% *dos dias nublados. Sabendo que na localidade em questão* 70% *dos dias são ensolarados, qual a porcentagem de acerto total do meteorologista?*

Sejam os eventos

$$\begin{aligned} A &= \{\textit{meteorologista acerta na previsão}\}, \\ B &= \{\textit{faz sol}\}. \end{aligned}$$

As informações que temos são as seguintes.

$$\mathbb{P}(A|B) = 0.80; \quad \mathbb{P}(A|B^c) = 0.60; \quad \mathbb{P}(B) = 0.70$$

Da regra da probabilidade total,

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c) \\ &= 0.80 \times 0.70 + 0.60 \times 0.30 = 0.56 + 0.18 = 0.74. \end{aligned}$$

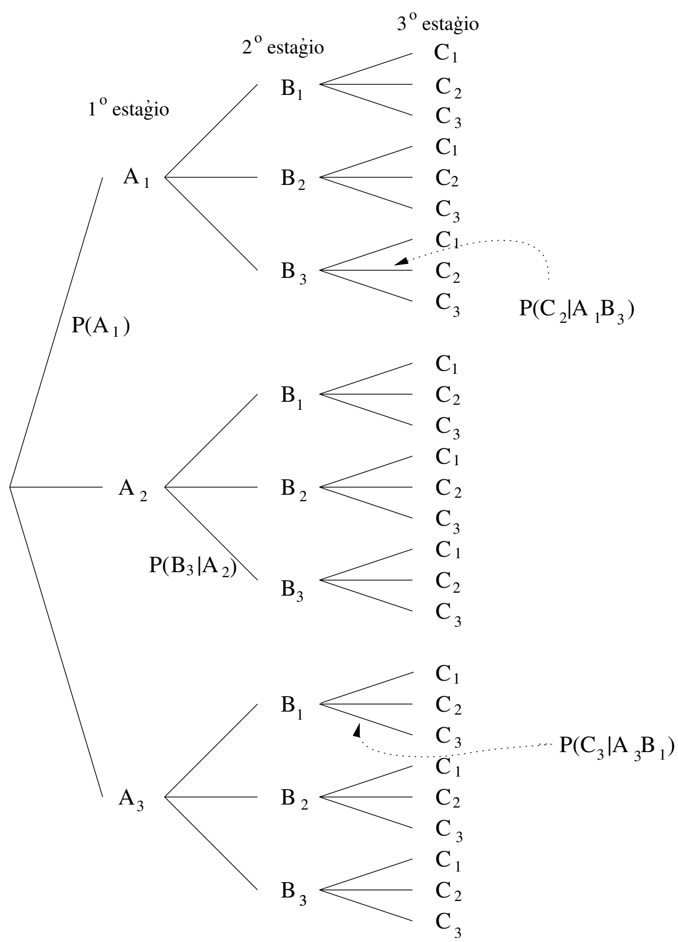

Figura 3.11 Árvore de probabilidades; apenas alguns ramos têm as respectivas probabilidades indicadas.

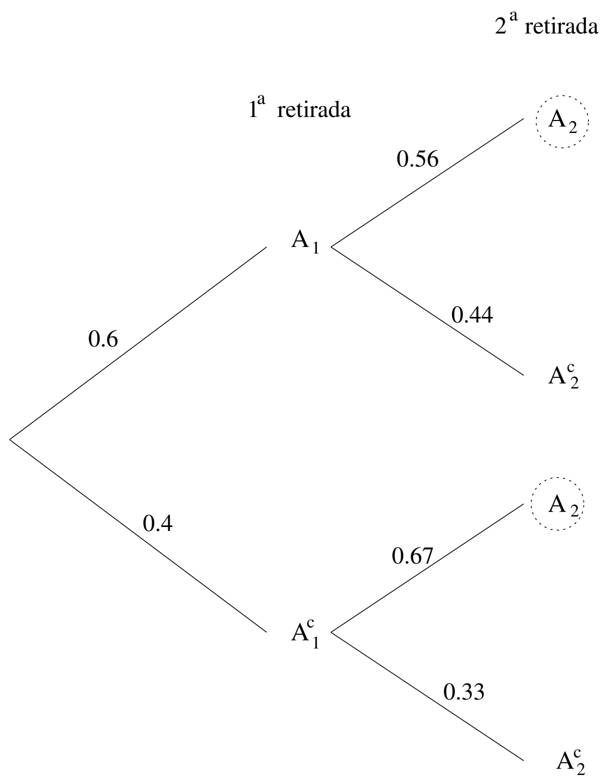

Figura 3.12 Árvore de probabilidades do Exemplo 3.18 com $K = 6$ e $M = 4$.

Regra de Bayes

Voltando ao contexto genérico do início da discussão sobre a regra da probabilidade total, suponha que o evento B ocorre. Qual é então a probabilidade de ocorrência de A_i, $i = 1, \ldots, n$?

Queremos determinar $\mathbb{P}(A_i|B)$. De (3.52) e (3.57),

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)}, \tag{3.66}$$

e da regra da probabilidade total, chegamos à *regra de Bayes*:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{n} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}. \tag{3.67}$$

Exemplo 3.20 *Uma vacina tem 90% de eficiência na imunização contra certa moléstia, que acomete 50% da população não vacinada. Suponha que, após uma campanha de vacinação em que 70% da população seja atingida, um paciente chegue a um hospital com a moléstia em questão, mas sem saber se tomou a vacina ou não. Qual é a probabilidade de que a tenha tomado?*

Sejam os eventos

$$\begin{aligned} A &= \{\textit{paciente foi acometido por moléstia}\}, \\ B &= \{\textit{paciente tomou vacina}\}. \end{aligned}$$

Queremos $\mathbb{P}(B|A)$. As informações que temos são as seguintes.

$$\mathbb{P}(A|B) = 0.10; \quad \mathbb{P}(A|B^c) = 0.50; \quad \mathbb{P}(B) = 0.70 \tag{3.68}$$

Da regra de Bayes (3.67),

$$\begin{aligned} \mathbb{P}(B|A) &= \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c)} \\ &= \frac{0.10 \times 0.70}{0.10 \times 0.70 + 0.50 \times 0.30} \\ &= \frac{0.07}{0.07 + 0.15} = \frac{0.07}{0.22} = 0.32. \end{aligned}$$

3.4.1 Independência

Dado um espaço de probabilidades $(\Omega, \mathcal{E}, \mathbb{P})$ e dois eventos $A, B \in \mathcal{E}$, dizemos que A é independente de B se

$$\mathbb{P}(A|B) = \mathbb{P}(A). \tag{3.69}$$

Pela regra do produto (3.57), (3.69) implica que

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \tag{3.70}$$

condição que implica (3.69), e logo é equivalente a (3.69). Podemos então tomá-la como condição de independência. Sua vantagem é que ela não envolve quocientes (o que evita a preocupação da divisão por 0), e é evidentemente simétrica, o que nos permite dizer que A e B são independentes (entre si) se satisfizerem (3.70) (ou (3.69)). (Mas (3.69) transmite de forma mais direta a idéia de independência.)

No Exemplo 3.1 (lançamento de dado equilibrado), sejam os eventos A, B como em (3.53,3.54). Então (3.55) nos diz que

$$\mathbb{P}(B|A) = \frac{2}{3} \neq \frac{1}{2} = \mathbb{P}(B),$$

e logo A e B *não* são independentes (neste caso, dizemos que são *dependentes*).

Mas sendo $B' =$\{o número lançado é maior do que 2\}, então

$$\begin{aligned} \mathbb{P}(B') &= \mathbb{P}(\{3,4,5,6\}) = \frac{4}{6} = \frac{2}{3}, \\ \mathbb{P}(A \cap B') &= \mathbb{P}(\{4,6\}) = \frac{2}{6} = \frac{1}{3}. \end{aligned}$$

Logo

$$\mathbb{P}(A \cap B') = \frac{1}{2}\frac{2}{3} = \mathbb{P}(A)\mathbb{P}(B'),$$

e A e B' são independentes.

A independência de dois eventos A e B se estende para os complementares.

Proposição 3.21 *Se A e B forem eventos independentes de dado espaço de probabilidades, então*

$$A \text{ e } B^c \text{ são independentes}, \tag{3.71}$$

$$A^c \text{ e } B \text{ são independentes}, \tag{3.72}$$

$$A^c \text{ e } B^c \text{ são independentes}. \tag{3.73}$$

Demonstração

$$\begin{aligned} (3.71): \quad \mathbb{P}(A \cap B^c) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(A)[1 - \mathbb{P}(B)] = \mathbb{P}(A)\mathbb{P}(B^c) \end{aligned} \tag{3.74}$$

Argumento similar para (3.72) e (3.73). □

Mais de dois eventos

Para $n \geq 3$ fixo, dizemos que os eventos $A_1, A_2, \dots, A_n$ de um espaço de probabilidades são *(mutuamente) independentes* se para todo $k = 2, \dots, n$ e $1 \leq i_1 < \ldots < i_k \leq n$, temos

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \ldots \mathbb{P}(A_{i_k}). \tag{3.75}$$

Isto é equivalente à seguinte condição. Dados dois subconjuntos disjuntos I e J quaisquer de índices em $\{1, \dots, n\}$, ambos não vazios, temos que os dois eventos

$$\cap_{i \in I} A_i \text{ e } \cap_{j \in J} A_j \text{ são independentes.}$$

Uma proposição semelhante à Proposição 3.21 vale para mais do que dois eventos (mutuamente) independentes, qual seja, a de que a independência é preservada se trocarmos qualquer subfamília de eventos por seus respectivos complementares.

Observação 3.22 (Amostragem) *Na amostragem casual simples com e sem reposição que discutimos na Seção 3.2, vamos considerar os eventos das sucessivas repetições dos sorteios que determinam a amostra.*

No primeiro caso, devido à reposição, os sorteios são sempre feitos na mesma população, sob as mesmas condições. Supomos (até o momento implicitamente) outras condições de independência, de forma que podemos dizer que os eventos de cada sorteio são mutuamente independentes.

Isto já não pode ser o caso quando não há reposição, pois as alterações que os resultados dos sucessivos sorteios vão produzindo na população induzem inevitavelmente dependência entre os eventos de diferentes sorteios.