Capítulo 5

Probabilidades

5.1 Introdução

Na primeira parte deste livro, vimos que a análise de um conjunto de dados por meio de técnicas numéricas e gráficas permite que tenhamos uma boa idéia da distribuição desse conjunto. Em particular, a distribuição de freqüências é um instrumento importante para avaliarmos a variabilidade das observações de um fenômeno aleatório. A partir dessas freqüências observadas podemos calcular medidas de posição e variabilidade, como média, mediana, desvio padrão etc. Essas freqüências e medidas calculadas a partir dos dados são *estimativas* de quantidades desconhecidas, associadas em geral a populações das quais os dados foram extraídos na forma de *amostras*. Em particular, as freqüências (relativas) são estimativas de *probabilidades* de ocorrências de certos eventos de interesse. Com suposições adequadas, e sem observarmos diretamente o fenômeno aleatório de interesse, podemos criar um *modelo teórico* que reproduza de maneira razoável a distribuição das freqüências, quando o fenômeno é observado diretamente. Tais modelos são chamados *modelos probabilísticos* e serão objeto de estudo neste capítulo e nos subseqüentes.

Exemplo 5.1. Queremos estudar as freqüências de ocorrências das faces de um dado. Um procedimento a adotar seria lançar o dado certo número de vezes, n, e depois contar o número n_i de vezes em que ocorre a face i, $i = 1, 2, ..., 6$. As proporções n_i/n determinam a distribuição de freqüências do experimento realizado. Lançando o dado um número $n'(n' \neq n)$ de vezes, teríamos outra distribuição de freqüências, mas com um padrão que esperamos ser muito próximo do anterior.

O modelo probabilístico pode ser construído por meio de premissas, como se segue.

Primeiro, observamos que só podem ocorrer seis faces; a segunda consideração que se faz é que o dado seja perfeitamente equilibrado, de modo a não favorecer alguma face em particular. Com essas suposições, cada face deve ocorrer o mesmo número de vezes quando o dado é lançado n vezes, e, portanto, a proporção de ocorrência de cada face deve ser 1/6. Nessas condições, o modelo teórico (ou probabilístico) para o experimento é dado na Tabela 5.1.

Tabela 5.1: Modelo para lançamento de um dado.

Face	1	2	3	4	5	6	Total
Freqüência teórica	1/6	1/6	1/6	1/6	1/6	1/6	1

Exemplo 5.2. De um grupo de duas mulheres (M) e três homens (H), uma pessoa será sorteada para presidir uma reunião. Queremos saber as probabilidades de o presidente ser do sexo masculino ou feminino. Observamos que: (i) só existem duas possibilidades: ou a pessoa sorteada é do sexo masculino (H) ou é do sexo feminino (M); (ii) supondo que o sorteio seja honesto e que cada pessoa tenha igual chance de ser sorteada, teremos o modelo probabilístico da Tabela 5.2 para o experimento.

Tabela 5.2: Modelo teórico para o Exemplo 5.2.

Sexo	*M*	*H*	Total
Freqüência teórica	2/5	3/5	1

Dos exemplos acima, verificamos que todo experimento ou fenômeno que envolva um elemento casual terá seu modelo probabilístico especificado quando estabelecermos:

(a) um *espaço amostral*, Ω, que consiste, no caso discreto, da enumeração (finita ou infinita) de todos os resultados possíveis do experimento em questão:

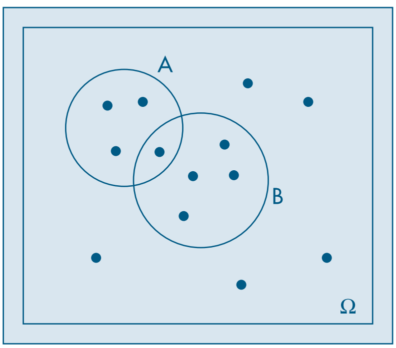
$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$$

(os elementos de Ω são os *pontos amostrais* ou *eventos elementares*);

(b) uma *probabilidade*, $P(\omega)$, para cada ponto amostral, de tal sorte que seja possível encontrar a probabilidade $P(A)$ de qualquer subconjunto A de Ω, isto é, a probabilidade do que chamaremos de um *evento aleatório* ou simplesmente *evento*.

Para ilustrar graficamente eventos, é costume utilizar-se os mesmos diagramas comumente usados na teoria dos conjuntos. Veja Morettin *et al.* (2005). Na Figura 5.1 ilustramos por um quadrado o espaço amostral, por círculos os eventos A e B e por pontos os pontos amostrais.

Figura 5.1: Espaço amostral e eventos aleatórios.



Exemplo 5.3. Lançamos uma moeda duas vezes. Se C indicar cara e R indicar coroa, então um espaço amostral será

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$$

onde $\omega_1 = (C, C)$, $\omega_2 = (C, R)$, $\omega_3 = (R, C)$, $\omega_4 = (R, R)$. É razoável supor que cada ponto ω_i tenha probabilidade 1/4, se a moeda for perfeitamente simétrica e homogênea.

Se designarmos por A o evento que consiste na obtenção de faces iguais nos dois lançamentos, então

$$P(A) = P\{\omega_1, \omega_4\} = 1/4 + 1/4 = 1/2.$$

De modo geral, se A for qualquer evento de Ω, então

$$P(A) = \sum_j P(\omega_j), \tag{5.1}$$

onde a soma é estendida a todos os pontos amostrais $\omega_j \in A$.

Exemplo 5.4. Uma fábrica produz determinado artigo. Da linha de produção são retirados três artigos, e cada um é classificado como bom (B) ou defeituoso (D). Um espaço amostral do experimento é

$$\Omega = \{BBB, BBD, BDB, DBB, DDB, DBD, BDD, DDD\}.$$

Se A designar o evento que consiste em obter dois artigos defeituosos, então $A = \{DDB, DBD, BDD\}$.

Exemplo 5.5. Considere o experimento que consiste em retirar uma lâmpada de um lote e medir seu "tempo de vida" antes de se queimar. Um espaço amostral conveniente é

$$\Omega = \{t \in \mathbb{R} : t \geqslant 0\},$$

isto é, o conjunto de todos os números reais não negativos. Se A indicar o evento "o tempo de vida da lâmpada é inferior a 20 horas", então $A = \{t : 0 \leqslant t < 20\}$. Esse é um exemplo de um espaço amostral *contínuo*, contrastado com os anteriores, que são *discretos*.

Problemas

1. Uma urna contém duas bolas brancas (B) e três bolas vermelhas (V). Retira-se uma bola ao acaso da urna. Se for branca, lança-se uma moeda; se for vermelha, ela é devolvida à urna e retira-se outra. Dê um espaço amostral para o experimento.

2. Lance um dado até que a face 5 apareça pela primeira vez. Enumere os possíveis resultados desse experimento.

3. Três jogadores A, B e C disputam um torneio de tênis. Inicialmente, A joga com B e o vencedor joga com C, e assim por diante. O torneio termina quando um jogador ganha duas vezes em seguida ou quando são disputadas, ao todo, quatro partidas. Quais são os resultados possíveis do torneio?

4. Duas moedas são lançadas. Dê dois possíveis espaços amostrais para esse experimento. Represente um deles como o produto cartesiano de dois outros espaços amostrais (ver Morettin *et al.*, 1999, para o conceito de produto cartesiano).

5. Uma moeda e um dado são lançados. Dê um espaço amostral do experimento e depois represente-o como produto cartesiano dos dois espaços amostrais, correspondente aos experimentos considerados individualmente.

6. Defina um espaço amostral para cada um dos seguintes experimentos aleatórios:
   (a) Lançamento de dois dados; anota-se a configuração obtida.
   (b) Numa linha de produção conta-se o número de peças defeituosas num intervalo de uma hora.
   (c) Investigam-se famílias com três crianças, anotando-se a configuração segundo o sexo.
   (d) Numa entrevista telefônica com 250 assinantes, anota-se se o proprietário tem ou não máquina de secar roupa.
   (e) Mede-se a duração de lâmpadas, deixando-as acesas até que se queimem.
   (f) Um fichário com dez nomes contém três nomes de mulheres. Seleciona-se ficha após ficha, até o último nome de mulher ser selecionado, e anota-se o número de fichas selecionadas.
   (g) Lança-se uma moeda até aparecer cara e anota-se o número de lançamentos.
   (h) Um relógio mecânico pode parar a qualquer momento por falha técnica. Mede-se o ângulo (em graus) que o ponteiro dos segundos forma com o eixo imaginário orientado do centro ao número 12.
   (i) Mesmo enunciado anterior, mas supondo que o relógio seja elétrico e, portanto, seu ponteiro dos segundos mova-se continuamente.
   (j) De um grupo de cinco pessoas $\{A, B, C, D, E\}$, sorteiam-se duas, uma após outra, com reposição, e anota-se a configuração formada.
   (l) Mesmo enunciado que *(j)*, sem reposição.
   (m) Mesmo enunciado que *(j)*, mas as duas selecionadas simultaneamente.
   (n) De cada família entrevistada numa pesquisa, anotam-se a classe social a que pertence (A, B, C, D) e o estado civil do chefe da família.

5.2 Algumas Propriedades

Sendo o modelo probabilístico um modelo teórico para as freqüências relativas, de suas propriedades podemos obter algumas das propriedades das probabilidades, que estudaremos a seguir.

Como a freqüência relativa é um número entre 0 e 1, temos que

$$0 < P(A) < 1, \tag{5.2}$$

para qualquer evento A. Será útil considerar o espaço todo Ω e o conjunto vazio $\emptyset$ como eventos. O primeiro é denominado *evento certo* e o segundo, *evento impossível*, e temos

$$P(\Omega) = 1, \; P(\emptyset) = 0. \tag{5.3}$$

Exemplo 5.6. Na Tabela 5.3 temos dados referentes a alunos matriculados em quatro cursos de uma universidade em dado ano.

Tabela 5.3: Distribuição de alunos segundo o sexo e escolha de curso.

Curso \ Sexo	Homens (H)	Mulheres (F)	Total
Matemática Pura (M)	70	40	110
Matemática Aplicada (A)	15	15	30
Estatística (E)	10	20	30
Computação (C)	20	10	30
Total	115	85	200

Vamos indicar por M o evento que ocorre quando, escolhendo-se ao acaso um aluno do conjunto desses quatro cursos, ele for um estudante de Matemática Pura. A, E, C, H e F têm significados análogos. Dessa maneira, vemos que $P(E) = 30/200$, ao passo que $P(H) = 115/200$.

Dados os eventos A e H, podemos considerar dois novos eventos:

- $A \cup H$, chamado a *reunião* de A e H, quando pelo menos um dos eventos ocorre;
- $A \cap H$, chamado a *intersecção* de A e H, quando A e H ocorrem simultaneamente.

É fácil ver que $P(A \cap H) = 15/200$, pois o aluno escolhido terá de estar, ao mesmo tempo, matriculado no curso de Matemática Aplicada e ser homem.

Vemos que $P(A) = 30/200$ e $P(H) = 115/200$; suponha que nosso cálculo para $P(A \cup H)$ fosse

$$P(A \cup H) = P(A) + P(H) = \frac{30}{200} + \frac{115}{200} = \frac{145}{200}.$$

Se assim o fizéssemos, estaríamos contando duas vezes os alunos que são homens e estão matriculados no curso de Matemática Aplicada, como destacado na Tabela 5.3. Portanto, a resposta correta é

$$P(A \cup H) = P(A) + P(H) - P(A \cap H) = \frac{30}{200} + \frac{115}{200} - \frac{15}{200} = \frac{130}{200}.$$

No entanto, considerando-se os eventos A e C, vemos que $P(A) = 30/200$, $P(C) = 30/200$ e $P(A \cup C) = 60/200 = P(A) + P(C)$. Nesse caso, os eventos A e C são disjuntos ou *mutuamente exclusivos*, pois se A ocorre, então C não ocorre e vice-versa. Aqui, $A \cap C = \emptyset$ e $P(A \cap C) = 0$.

Portanto, se U e V são dois eventos quaisquer, teremos a chamada *regra da adição de probabilidades*

$$P(U \cup V) = P(U) + P(V) - P(U \cap V), \tag{5.4}$$

que se reduz a

$$P(U \cup V) = P(U) + P(V), \tag{5.5}$$

se U e V são eventos mutuamente exclusivos. Veja o Problema 58.

Suponha, agora, que estejamos somente interessados em saber se um estudante escolhido ao acaso está matriculado como aluno de Matemática Pura, Aplicada, Estatística ou Computação, não interessando saber se é homem ou mulher. Seja $B = M \cup E \cup C$. Então $A \cup B = \Omega$ e $A \cap B = \emptyset$. Dizemos que A e B são *complementares* e $P(A) = 30/200$, $P(B) = 110/200 + 30/200 + 30/200 = 170/200$, isto é, $P(A) + P(B) = 1$.

De modo geral, vamos indicar por A^c o complementar de um evento qualquer A, e teremos então

$$P(A) + P(A^c) = 1. \tag{5.6}$$

As operações de reunião, intersecção e complementação entre eventos possuem propriedades análogas àquelas válidas para operações entre conjuntos. Ver Morettin *et al.* (2005). Por exemplo:

(a) $(A \cap B)^c = A^c \cup B^c$
(b) $(A \cup B)^c = A^c \cap B^c$
(c) $A \cap \emptyset = \emptyset$, $A \cap \Omega = A$
(d) $\emptyset^c = \Omega$, $\Omega^c = \emptyset$
(e) $A \cap A^c = \emptyset$
(f) $A \cup A^c = \Omega$
(g) $A \cup \emptyset = A$, $A \cup \Omega = \Omega$
(h) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Vejamos um exemplo de aplicação das propriedades das probabilidades.

Exemplo 5.7. Consideremos um experimento aleatório e os eventos A e B associados, tais que $P(A) = 1/2$, $P(B) = 1/3$ e $P(A \cap B) = 1/4$. Então temos:

(a) $P(A^c) = 1 - P(A) = 1 - 1/2 = 1/2$;
$P(B^c) = 1 - P(B) = 1 - 1/3 = 2/3$.

(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1/2 + 1/3 - 1/4 = 7/12$.

(c) $P(A^c \cap B^c) = P[(A \cup B)^c] = 1 - P(A \cup B) = 1 - 7/12 = 5/12$.

(d) $P(A^c \cup B^c) = P[(A \cap B)^c] = 1 - P(A \cap B) = 1 - 1/4 = 3/4$.

(e) Calculemos $P(A^c \cap B)$, isto é, a probabilidade de que ocorra B e não ocorra A. Podemos escrever

$$B = (A \cap B) \cup (A^c \cap B),$$

ou seja, B pode ocorrer com A ou (exclusivo) com A^c. Logo,

$$P(B) = P(A \cap B) + P(A^c \cap B),$$

do que decorre

$$P(A^c \cap B) = P(B) - P(A \cap B) = 1/3 - 1/4 = 1/12.$$

Consideremos, agora, uma situação historicamente importante, a saber, aquela em que temos um espaço amostral finito, $\Omega = \{\omega_1, ..., \omega_n\}$, em que todos os pontos têm a mesma probabilidade $1/n$. Se A for um evento contendo m pontos amostrais, então

$$P(A) = \frac{m}{n}.$$

Nesse caso, não é necessário explicitar completamente Ω e A, bastando calcular m e n, chamados, respectivamente, *número de casos favoráveis* e *número de casos possíveis*. Para tanto, são usados os métodos clássicos de contagem da análise combinatória. Um princípio fundamental de contagem nos diz que, se uma tarefa pode ser executada em duas etapas, a primeira podendo ser realizada de p maneiras e a segunda de q maneiras, então as duas podem ser realizadas simultaneamente de pq maneiras. Esse é o chamado *princípio multiplicativo*.

Exemplo 5.8. Suponha que num lote com 20 peças existam cinco defeituosas. Escolhemos quatro peças do lote ao acaso, ou seja, uma *amostra* de quatro elementos, de modo que a ordem dos elementos seja irrelevante.

Dessa maneira, o número de amostras com quatro elementos que podemos extrair do lote é $\binom{20}{4}$, ou seja, combinações de 20 elementos, tomados quatro a quatro. Suponha que queiramos calcular a probabilidade de se escolher duas defeituosas na amostra. Pelo visto acima, $\binom{20}{4}$ é o número de pontos do espaço amostral. Seja A o evento que consiste em escolher duas defeituosas na amostra. Segue-se que $m = \binom{5}{2}\binom{15}{2}$, pois podemos escolher na amostra de quatro elementos duas defeituosas e duas não-defeituosas simultaneamente de $\binom{5}{2}\binom{15}{2}$ maneiras, usando o princípio multiplicativo. Logo,

$$P(A) = \frac{\binom{5}{2}\binom{15}{2}}{\binom{20}{4}} = 0{,}217.$$

Exemplo 5.9. O jogo da Megasena consiste em escolher 6 dezenas dentre as 60 dezenas (01, 02, ..., 59, 60). O jogador pode marcar num cartão de 6 a 15 dezenas. Os custos (em reais) de cada jogo estão relacionados abaixo.

Dezenas	Custo
6	1,00
7	7,00
8	28,00
9	84,00
10	210,00
11	462,00
12	924,00
13	1.716,00
14	3.005,00
15	5.005,00

Temos, ao todo, $\binom{60}{6} = 50.063.860$ possibilidades. Portanto, com um jogo único de R$ 1,00 (seis dezenas), a probabilidade de ganhar o prêmio máximo é $1/\binom{60}{6}$, ou seja, aproximadamente, uma chance em 50 milhões. Por quê o jogo com 7 dezenas custa R$ 7,00? Porque com 7 dezenas podemos formar $\binom{7}{6} = 7$ jogos de 6 dezenas. Ou seja, fazer um jogo com

7 dezenas ou 7 jogos com 6 dezenas são ações equivalentes, em termos de probabilidade de ganhar. Do mesmo modo, um jogo de 15 dezenas custa R$ 5.005,00, porque com 15 dezenas podemos formar $\binom{15}{6} = 5.005$ jogos de 6 dezenas. Portanto, é mais fácil preencher um boleto com 15 dezenas do que 5.005 boletos com 6 dezenas, já que as probabilidades associadas são iguais.

Problemas

7. No Problema 4, liste os eventos:
   (*a*) pelo menos uma cara;
   (*b*) duas caras;
   (*c*) o complementar do evento em (*b*).

8. Expresse em termos de operações entre eventos:
   (*a*) A ocorre mas B não ocorre;
   (*b*) exatamente um dos eventos A e B ocorre;
   (*c*) nenhum dos dois eventos A e B ocorre.

9. No espaço amostral do Problema 3, atribua a cada ponto contendo k letras a probabilidade $1/2^k$ (assim, AA tem probabilidade 1/4).
   (*a*) Mostre que a soma das probabilidades dos pontos do espaço amostral é 1.
   (*b*) Calcule a probabilidade de que A vença (um jogador vence quando ganha duas partidas seguidas). Em seguida, calcule a probabilidade de que B vença.
   (*c*) Qual a probabilidade de que não haja decisão?

10. No Problema 2, suponha que 5 indique o aparecimento da face 5 e Q indique que apareceu outra face qualquer diferente da 5. Atribua probabilidade $(5/6)^k(1/6)$ a cada ponto com k letras iguais a Q seguidas de 5.
   (*a*) Mostre que a soma das probabilidades dos pontos amostrais é igual a um (aqui você deve usar o resultado da soma dos termos de uma seqüência geométrica infinita).
   (*b*) Calcule a probabilidade de que a face 5 apareça após três lançamentos do dado.

11. Dentre seis números positivos e oito negativos, dois números são escolhidos ao acaso (sem reposição) e multiplicados. Qual a probabilidade de que o produto seja positivo?

12. Considere o lançamento de dois dados. Considere os eventos: A = soma dos números obtidos igual a 9, e B = número no primeiro dado maior ou igual a 4. Enumere os elementos de A e B. Obtenha $A \cup B$, $A \cap B$ e A^c.

13. Obtenha as probabilidades dos eventos que aparecem nos Problemas 7 e 12.

14. Que suposições devem ser feitas para que os resultados dos experimentos abaixo possam ser considerados eqüiprováveis?
   (*a*) Lançamento de um dado.
   (*b*) Opinião de moradores de uma cidade sobre um projeto governamental.
   (*c*) Preço de uma ação no fim da próxima semana.