Pedro A. Morettin
Wilton de O. Bussab

ESTATÍSTICA BÁSICA

6ª edição
Revista e atualizada

Editora Saraiva

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:

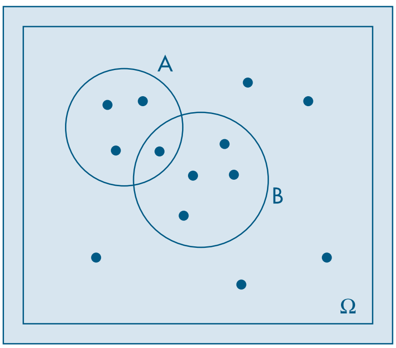
$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$$

(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 $\varnothing$ como eventos. O primeiro é denominado *evento certo* e o segundo, *evento impossível*, e temos

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

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, $A \cap C = \varnothing$ 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, $\Omega = \{\omega_1, \ldots, \omega_n\}$, 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.

5.3 Probabilidade Condicional e Independência

Voltemos à Tabela 5.3 do Exemplo 5.6. Dado que um estudante, escolhido ao acaso, esteja matriculado no curso de Estatística, a probabilidade de que seja mulher é 20/30 = 2/3. Isso porque, do total de 30 alunos que estudam Estatística, 20 são mulheres. Escrevemos

$$P(\text{mulher}\,|\,\text{Estatística}) = \frac{2}{3}.$$

Para dois eventos quaisquer A e B, sendo $P(B) > 0$, definimos a *probabilidade condicional de* A *dado* B, $P(A|B)$, como sendo

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{5.7}$$

Para o exemplo mencionado, se B e A indicam, respectivamente, os eventos "aluno matriculado em Estatística" e "aluno é mulher", então

$$P(A|B) = \frac{20/200}{30/200} = \frac{2}{3},$$

como havíamos obtido.

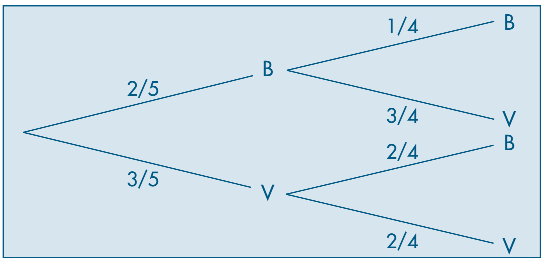
Observe que $P(A) = P(\text{mulher}) = 85/200 = 17/40$, e com a informação de que B ocorreu (o aluno é matriculado em Estatística), obtemos $P(A|B) = 2/3$. Podemos dizer que $P(A)$ é a probabilidade *a priori* de A e, com a informação adicional de que B ocorreu, obtemos a probabilidade *a posteriori* $P(A|B)$. Note que, nesse caso, $P(A|B) > P(A)$, logo a informação de que B ocorreu aumentou a chance de A ocorrer.

Da relação (5.7) obtemos a chamada *regra do produto de probabilidades*,

$$P(A \cap B) = P(B)\, P(A|B). \tag{5.8}$$

Exemplo 5.10. Uma urna contém duas bolas brancas (B) e três vermelhas (V). Suponha que são sorteadas duas bolas ao acaso, *sem reposição*. Isso significa que escolhemos a primeira bola, verificamos sua cor e não a devolvemos à urna; misturamos as bolas restantes e retiramos a segunda. O diagrama em árvore da Figura 5.2 ilustra as possibilidades. Em cada "galho" da árvore estão indicadas as probabilidades de ocorrência, sendo que para as segundas bolas as probabilidades são condicionais. A probabilidade do resultado conjunto é dada, então, por (5.8). Veja a Tabela 5.4.

Figura 5.2: Diagrama em árvore para a extração de duas bolas de uma urna, sem reposição.



Se A indicar o evento “bola branca na segunda extração”, então

$$P(A) = P(BB) + P(VB) = \frac{2}{20} + \frac{6}{20} = \frac{2}{5}.$$

Tabela 5.4: Resultados e probabilidades para o experimento do Exemplo 5.10.

Resultados	Probabilidades
BB	$2/5 \times 1/4 = 2/20$
BV	$2/5 \times 3/4 = 6/20$
VB	$3/5 \times 2/4 = 6/20$
VV	$3/5 \times 2/4 = 6/20$
Total	1

Exemplo 5.11. Imagine, agora, que as duas extrações são feitas da mesma urna do exemplo anterior, mas a primeira bola é *reposta* na urna antes da extração da segunda. Nessas condições, as extrações são independentes, pois o resultado de uma extração não tem influência no resultado da outra. Obtemos a situação da Figura 5.3 e da Tabela 5.5.

Figura 5.3: Diagrama em árvore para a extração de duas bolas de uma urna, com reposição.

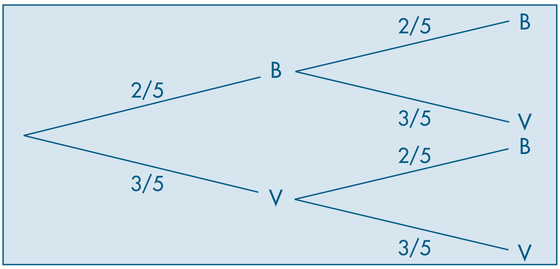


Tabela 5.5: Resultados e probabilidades para o experimento do Exemplo 5.11.

Resultados	Probabilidades
BB	$2/5 \times 2/5 = 4/25$
BV	$2/5 \times 3/5 = 6/25$
VB	$3/5 \times 2/5 = 6/25$
VV	$3/5 \times 3/5 = 9/25$
Total	1

Observe que, aqui,

$$P(\text{branca na } 2^{\underline{a}} \mid \text{branca na } 1^{\underline{a}}) = 2/5 = P(\text{branca na } 2^{\underline{a}}),$$

ou seja, se indicarmos por A e B os eventos "bola branca na segunda extração" e "bola branca na primeira extração", respectivamente, então $P(A|B) = P(A)$. Nesse caso, dizemos que o evento A *independe* do evento B e, usando (5.8), temos

$$P(A \cap B) = P(A)\, P(B). \tag{5.9}$$

É fácil ver que se A independe de B, então B independe de A — dizemos que A e B são independentes. A fórmula (5.9) pode ser tomada como definição de independência entre dois eventos, ou seja, A e B são *independentes* se, e somente se, (5.9) for válida.

Exemplo 5.12. Considere ainda a urna dos dois exemplos anteriores, mas vamos fazer três extrações *sem reposição*. Indiquemos por V_i ou B_i a obtenção de bola vermelha ou branca na i-ésima extração, respectivamente, $i = 1, 2, 3$. Obtemos a Figura 5.4 e a Tabela 5.6.

Figura 5.4: Diagrama em árvore para a extração de três bolas de uma urna, sem reposição.

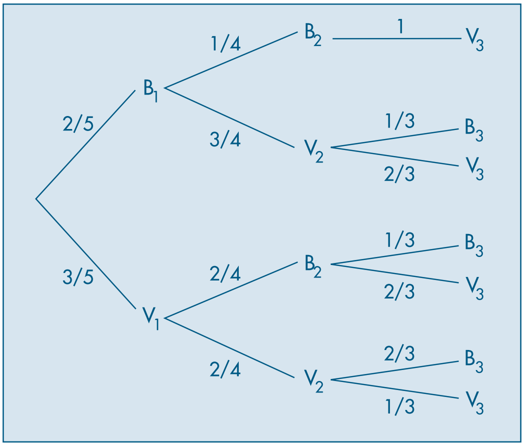


Tabela 5.6: Resultados e probabilidades para o experimento do Exemplo 5.12.

Resultados	Probabilidades
$B_1B_2V_3$	$2/5 \times 1/4 \times 1 = 2/20 = 6/60$
$B_1V_2B_3$	$2/5 \times 3/4 \times 1/3 = 6/60$
$B_1V_2V_3$	$2/5 \times 3/4 \times 2/3 = 12/60$
$V_1B_2B_3$	$3/5 \times 2/4 \times 1/3 = 6/60$
$V_1B_2V_3$	$3/5 \times 2/4 \times 2/3 = 12/60$
$V_1V_2B_3$	$3/5 \times 2/4 \times 2/3 = 12/60$
$V_1V_2V_3$	$3/5 \times 2/4 \times 1/3 = 6/60$
Total	$60/60 = 1$

Observe que $P(B_2|B_1) = 1/4$, ao passo que $P(V_3|B_1 \cap B_2) = 1$; daí,

$$P(B_1 \cap B_2 \cap V_3) = P(B_1)\, P(B_2|B_1)\, P(V_3|B_1 \cap B_2) = 2/5 \times 1/4 \times 1 = 1/10.$$

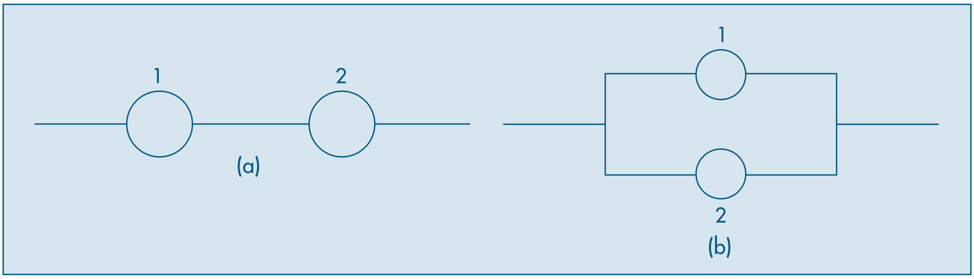
De modo geral, dados três eventos A, B e C, temos que

$$P(A \cap B \cap C) = P(A)\, P(B|A)\, P(C|A \cap B). \tag{5.10}$$

Essa relação pode ser estendida para um número finito qualquer de eventos. Veja o Problema 60.

Exemplo 5.13. A *teoria da confiabilidade* estuda sistemas e seus componentes, como por exemplo sistemas mecânicos e eletrônicos (um automóvel ou um computador) e sistemas biológicos, como o corpo humano. O objetivo da teoria é estudar as relações entre o funcionamento dos componentes e do sistema. A Figura 5.5 (a) ilustra um sistema composto de dois componentes ligados em *série*.

Figura 5.5: Sistema com dois componentes (a) em série (b) em paralelo.



O sistema da figura funcionará se os componentes 1 e 2 funcionarem simultaneamente. Se um dos componentes falhar, o sistema também falhará. Supondo que os componentes funcionem *independentemente*, e se p_i for a probabilidade de o componente i (i = 1,2) funcionar, então a probabilidade de o sistema funcionar será

$$P(F) = P(A_1 \cap A_2) = P(A_1)P(A_2) = p_1 p_2,$$

onde indicamos por F o evento "o sistema funciona" e por A_i o evento "o componente i funciona", i = 1, 2.

A probabilidade p_i é a chamada *confiabilidade do componente* i e $P(F) = h(p_1, p_2) = p_1 p_2$ a *confiabilidade do sistema*.

Se os componentes 1 e 2 estiverem em *paralelo*, como na Figura 5.5 (b), então o sistema funcionará se *pelo menos um* dos dois componentes funcionar. Ou seja,

$$P(F) = P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = p_1 + p_2 - p_1 p_2$$

e a confiabilidade do sistema é $h(p_1, p_2) = p_1 + p_2 - p_1 p_2$.

Vejamos agora o conceito de independência para três eventos: dizemos que os eventos A, B e C são *independentes* se, e somente se,

$$\begin{aligned} P(A \cap B) &= P(A)\, P(B), \\ P(A \cap C) &= P(A)\, P(C), \\ P(B \cap C) &= P(B)\, P(C), \\ P(A \cap B \cap C) &= P(A)\, P(B)\, P(C). \end{aligned} \tag{5.11}$$

Se apenas as três primeiras relações de (5.11) estiverem satisfeitas, dizemos que os eventos A, B e C são *mutuamente independentes*. É possível que três eventos sejam mutuamente independentes, mas não sejam completamente independentes. Veja o Problema 59.

A definição pode ser estendida facilmente para um número finito qualquer de eventos. Veja o Problema 61.

Problemas

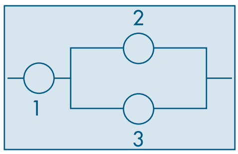
15. Considere uma urna contendo três bolas pretas e cinco bolas vermelhas. Retire duas bolas da urna, sem reposição.
    (a) Obtenha os resultados possíveis e as respectivas probabilidades.
    (b) Mesmo problema, para extrações com reposição.

16. No problema anterior, calcule as probabilidades dos eventos:
    (a) Bola preta na primeira e segunda extrações.
    (b) Bola preta na segunda extração.
    (c) Bola vermelha na primeira extração.

17. A probabilidade de que A resolva um problema é de 2/3, e a probabilidade de que B o resolva é de 3/4. Se ambos tentarem independentemente, qual a probabilidade de o problema ser resolvido?

18. Um dado é viciado, de tal forma que a probabilidade de sair um certo ponto é proporcional ao seu valor (por exemplo, o ponto 6 é três vezes mais provável de sair do que o ponto 2). Calcular:
    (a) a probabilidade de sair 5, sabendo-se que o ponto que saiu é ímpar;
    (b) a probabilidade de tirar um número par, sabendo-se que saiu um número maior que 3.

19. As probabilidades de que dois eventos independentes ocorram são p e q, respectivamente. Qual a probabilidade:
    (a) de que nenhum desses eventos ocorra?
    (b) de que pelo menos um desses eventos ocorra?



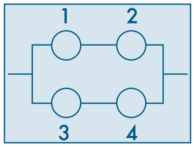
20. Na figura ao lado temos um sistema com três componentes funcionando independentemente, com confiabilidades p_1, p_2 e p_3. Obtenha a confiabilidade do sistema.

21. Na tabela abaixo, os números que aparecem são probabilidades relacionadas com a ocorrência de A, B, $A \cap B$ etc. Assim, $P(A) = 0{,}10$, enquanto $P(A \cap B) = 0{,}04$.

	B	B^c	Total
A	0,04	0,06	0,10
A^c	0,08	0,82	0,90
Total	0,12	0,88	1,00

Verifique se A e B são independentes.

22. Supondo que todos os componentes do sistema da figura ao lado tenham a mesma confiabilidade p e funcionem independentemente, obtenha a confiabilidade do sistema.



5.4 O Teorema de Bayes

Uma das relações mais importantes envolvendo probabilidades condicionais é dada pelo Teorema de Bayes. A versão mais simples desse teorema é dada pela fórmula (5.12):

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B|A)}{P(B)}. \tag{5.12}$$

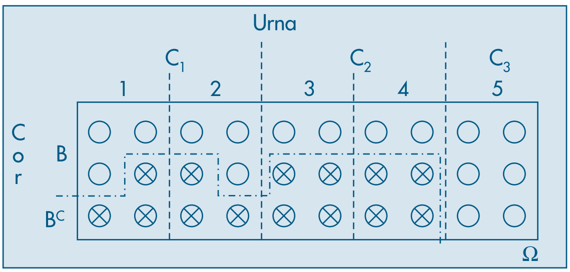
Como salientamos na seção anterior, temos a probabilidade inicial $P(A)$ e, dada a informação de que B ocorreu (ou dada a suposição de que B venha a ocorrer), obtemos a probabilidade *a posteriori* $P(A|B)$, dada por (5.12). Ou seja, *atualizamos* a probabilidade inicial, multiplicando-a por $\frac{P(B|A)}{P(B)}$. Observe que $P(A|B) > P(A)$ se $P(B|A) > P(B)$.

A forma geral do Teorema de Bayes será introduzida por um exemplo.

Exemplo 5.14. Temos cinco urnas, cada uma com seis bolas. Duas dessas urnas (tipo C_1) têm 3 bolas brancas, duas outras (tipo C_2) têm 2 bolas brancas, e a última urna (tipo C_3) tem 6 bolas brancas. Escolhemos uma urna ao acaso e dela retiramos uma bola. Qual a probabilidade de a urna escolhida ser do tipo C_3, sabendo-se que a bola sorteada é branca?

Na Figura 5.6 temos esquematizados o espaço amostral e os eventos de interesse.

Figura 5.6: Espaço amostral e eventos para o Exemplo 5.14.



Queremos encontrar $P(C_3|B)$, sabendo que

$$P(C_1) = 2/5,\quad P(B|C_1) = 1/2,$$
$$P(C_2) = 2/5,\quad P(B|C_2) = 1/3,$$
$$P(C_3) = 1/5,\quad P(B|C_3) = 1.$$

Da definição de probabilidade condicional, temos

$$P(C_3|B) = \frac{P(C_3 \cap B)}{P(B)} = \frac{P(C_3)P(B|C_3)}{P(B)}. \tag{5.13}$$

A segunda igualdade é devida à fórmula (5.8).

Precisamos encontrar o valor de $P(B)$, já que o numerador é conhecido. Como C_1, C_2 e C_3 são eventos mutuamente exclusivos, e reunidos formam o espaço amostral completo, podemos decompor o evento B na reunião de três outros, também mutuamente exclusivos, como segue (ver também a Figura 5.6):

$$B = (C_1 \cap B) \cup (C_2 \cap B) \cup (C_3 \cap B), \tag{5.14}$$

e então

$$\begin{aligned} P(B) &= P(C_1 \cap B) + P(C_2 \cap B) + P(C_3 \cap B) \\ &= P(C_1)\, P(B|C_1) + P(C_2)\, P(B|C_2) + P(C_3)\, P(B|C_3) \\ &= \frac{2}{5} \times \frac{1}{2} + \frac{2}{5} \times \frac{1}{3} + \frac{1}{5} \times 1 = \frac{8}{15}. \end{aligned}$$

Substituindo esse resultado em (5.13), obtemos

$$P(C_3|B) = \frac{1/5 \times 1}{8/15} = \frac{3}{8}.$$

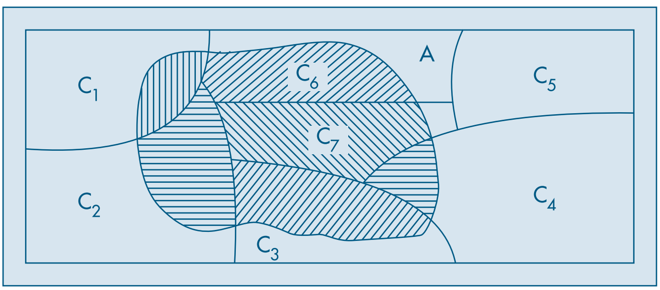
Podemos, agora, generalizar os resultados acima do seguinte modo: seja $\{C_1, C_2, ..., C_n\}$ uma partição do espaço amostral Ω, isto é,

$$C_i \cap C_j = \varnothing, \quad \text{sempre que} \quad i \neq j,$$
$$C_1 \cup C_2 \cup ... \cup C_n = \Omega.$$

Considere um evento qualquer A em Ω. Supomos conhecidas as probabilidades $P(C_i)$ e $P(A|C_i)$, $i = 1, 2, ..., n$.

Então, temos o seguinte resultado, ilustrado pela Figura 5.7.

Figura 5.7: Partição de um espaço amostral.



Teorema 5.1 (Bayes). A probabilidade de ocorrência do evento C_i, supondo-se a ocorrência do evento A, é dada por

$$P(C_i|A) = \frac{P(C_i)P(A|C_i)}{\sum_{j=1}^{n} P(C_j)P(A|C_j)}, \tag{5.15}$$

para todo $i = 1, 2, ..., n$.

Podemos pensar $C_1, ..., C_n$ como um conjunto de *hipóteses*, sendo somente uma delas verdadeira. Dado que A ocorreu, a probabilidade inicial de C_i, $P(C_i)$, é modificada de modo a se obter $P(C_i|A)$, dada por (5.15). Passamos da probabilidade *a priori* $P(C_i)$ para a probabilidade *a posteriori* $P(C_i|A)$, multiplicando a primeira por

$$\frac{P(A|C_i)}{\sum_{j=1}^{n} P(C_j)P(A|C_j)}. \tag{5.16}$$

Para A fixado, as probabilidades $P(A|C_i)$ em (5.15) são denominadas *verossimilhanças* das hipóteses $C_1, C_2, ..., C_n$. Vemos que $P(C_i|A) > P(C_i)$ se (5.16) for maior do que um, isto é, se $P(A|C_i) > P(A)$, onde $P(A)$ é o denominador de (5.16). Observe que esse denominador é uma média ponderada dos $P(A|C_j)$ e os pesos são as probabilidades $P(C_j)$, que têm soma unitária. Como o numerador é sempre uma das parcelas do denominador $P(A)$, torna-se indispensável o uso de um novo índice, j, na decomposição deste.

Exemplo 5.15. Para selecionar seus funcionários, uma empresa oferece aos candidatos um curso de treinamento durante uma semana. No final do curso, eles são submetidos a uma prova e 25% são classificados como bons (B), 50% como médios (M) e os restantes 25% como fracos (F). Para facilitar a seleção, a empresa pretende substituir o treinamento por um teste contendo questões referentes a conhecimentos gerais e específicos. Para isso, gostaria de conhecer qual a probabilidade de um indivíduo aprovado no teste ser considerado fraco, caso fizesse o curso. Assim, neste ano, antes do início do curso, os candidatos foram submetidos ao teste e receberam o conceito aprovado (A) ou reprovado (R). No final do curso, obtiveram-se as seguintes probabilidades condicionais:

$$P(A|B) = 0{,}80, \qquad P(A|M) = 0{,}50, \qquad P(A|F) = 0{,}20.$$

Queremos encontrar $P(F|A)$ e, pelo Teorema de Bayes, essa probabilidade é dada por

$$P(F|A) = \frac{P(A|F)P(F)}{P(A|B)P(B) + P(A|M)P(M) + P(A|F)P(F)}$$

$$= \frac{(0{,}20)(0{,}25)}{(0{,}80)(0{,}25) + (0{,}50)(0{,}50) + (0{,}20)(0{,}25)} = 0{,}10.$$

Então, apenas 10% dos aprovados é que seriam classificados como fracos durante o curso. De modo análogo podemos encontrar $P(B|A) = 0{,}40$ e $P(M|A) = 0{,}50$, que poderiam fornecer subsídios para ajudar na decisão de substituir o treinamento pelo teste.

Um gráfico em árvore pode ajudar bastante na solução de um problema envolvendo o Teorema de Bayes. Desse modo, para o Exemplo 5.15, teremos a Figura 5.8 e a Tabela 5.7. Assim, o numerador de $P(F|A)$ está assinalado com um pequeno círculo, ao passo que o denominador é a soma das três parcelas assinaladas com asterisco.

Figura 5.8: Diagrama em árvore para o Exemplo 5.15.

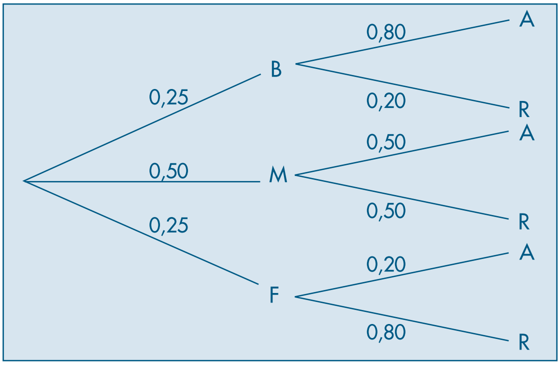


Tabela 5.7: Resultados e probabilidades para o Exemplo 5.15.

Resultados	Probabilidades
BA	(0,25) (0,80) = 0,20 *
BR	(0,25) (0,20) = 0,05
MA	(0,50) (0,50) = 0,25 *
MR	(0,50) (0,50) = 0,25
FA	(0,25) (0,20) = 0,05 * °
FR	(0,25) (0,80) = 0,20

O Teorema de Bayes, que aparentemente poderia ser encarado como mais um resultado na teoria de probabilidades, tem importância fundamental, pois fornece a base para uma abordagem da inferência estatística conhecida como *inferência bayesiana*. Esse ponto será abordado brevemente no Capítulo 11.

O Teorema de Bayes fornece um mecanismo formal para atualizar probabilidades, como já vimos acima. Vejamos mais um exemplo para ilustrar esse ponto.

Exemplo 5.16. A administração de um fundo de investimentos em ações pretende divulgar, após o encerramento do pregão, a probabilidade de queda de um índice da bolsa no dia seguinte, baseando-se nas informações disponíveis até aquele momento. Suponha que a previsão inicial seja de 0,10. Após encerrado o pregão, nova informação sugere uma alta do dólar frente ao real. A experiência passada indica que, quando houve queda da bolsa no dia seguinte, 20% das vezes foram precedidas por esse tipo de notícia, enquanto, nos dias em que a bolsa esteve em alta, apenas em 5% das vezes houve esse tipo de notícia no dia anterior.

Chamando de E o evento que indica "queda da bolsa", a sua probabilidade *a priori* é $P(E) = 0{,}10$, enquanto a probabilidade de alta é $P(E^c) = 0{,}90$. Se B indicar "alta do dólar", então as verossimilhanças são dadas por

$$P(B|E) = 0{,}20, \qquad P(B|E^c) = 0{,}05.$$

Logo, pelo Teorema de Bayes, teremos que

$$P(E|B) = \frac{P(E)\,P(B|E)}{P(E)P(B|E) + P(E^c)P(B|E^c)},$$

ou seja,

$$P(E|B) = \frac{(0{,}10)(0{,}20)}{(0{,}10)(0{,}20) + (0{,}90)(0{,}05)} = \frac{0{,}02}{0{,}065} = \frac{4}{13} = 0{,}31.$$

Portanto, a nova informação aumenta a probabilidade de que haja queda na bolsa de 10% para 31%.

Suponha, agora, que horas depois surja nova informação relevante: o Banco Central irá reduzir a taxa de juros vigente a partir do dia seguinte. Denotando-se, agora, por B_1 o evento "alta do dólar" e por B_2 o evento "queda na taxa de juros", o interesse será saber como essa nova informação, B_2, afetará a probabilidade calculada, $P(E|B_1)$. Segue-se que essa é agora a probabilidade *a priori* para E com respeito a B_2.

Novamente, informações passadas mostram que, dado que tenha havido alta do dólar e queda da bolsa, 10% das vezes foram precedidas por notícias de queda de juros, enquanto, dado que tenha havido alta do dólar e alta da bolsa, 60% das vezes foram precedidas de queda dos juros. Então, as verossimilhanças agora serão dadas por

$$P(B_2|E, B_1) = 0{,}10, \qquad P(B_2|E^c, B_1) = 0{,}60.$$

O Teorema de Bayes fica escrito agora na forma

$$P(E|B_1, B_2) = \frac{P(E|B_1)\,P(B_2|E, B_1)}{P(E|B_1)\,P(B_2|E, B_1) + P(E^c|B_1)\,P(B_2|E^c, B_1)},$$

do que segue que

$$P(E|B_1, B_2) = \frac{(0{,}31)(0{,}10)}{(0{,}31)(0{,}10) + (0{,}69)(0{,}60)} = \frac{0{,}031}{0{,}445} = 0{,}07.$$

Ou seja, a informação B_2 causa um decréscimo na probabilidade de queda da bolsa, de 0,31 para 0,07, que é menor ainda do que a probabilidade *a priori* inicial, P(E) = 0,10.

Observe que a probabilidade $P(E|B_1, B_2)$ pode ser escrita também como $P(E|B_1 \cap B_2)$, ou seja, temos a ocorrência simultânea dos eventos B_1 e B_2.

Problemas

23. Uma companhia produz circuitos em três fábricas, I, II e III. A fábrica I produz 40% dos circuitos, enquanto a II e a III produzem 30% cada uma. As probabilidades de que um circuito integrado produzido por essas fábricas não funcione são 0,01, 0,04 e 0,03, respectivamente. Escolhido um circuito da produção conjunta das três fábricas, qual a probabilidade de o mesmo não funcionar?

24. Considere a situação do problema anterior, mas suponha agora que um circuito escolhido ao acaso seja defeituoso. Determine qual a probabilidade de ele ter sido fabricado por I.

25. A urna I contém duas bolas pretas e três brancas, ao passo que a urna II contém três bolas pretas e três brancas. Escolhemos uma urna ao acaso e dela extraímos uma bola que tem cor branca. Se a bola é recolocada na urna, qual é a probabilidade de se retirar novamente uma bola branca da mesma urna?

5.5 Probabilidades Subjetivas

Na seção 5.1 vimos como associar probabilidades a eventos. Utilizamos um enfoque chamado *freqüentista*, pois se baseia na estabilidade das freqüências relativas e no fato de podermos, hipoteticamente, repetir um experimento várias vezes. Mas é óbvio que nem sempre podemos considerar replicações. Suponha que queiramos calcular a probabilidade de chover no dia 12 de janeiro do próximo ano, na cidade de São Paulo. Evidentemente, se considerarmos o evento A = chover em São Paulo no dia 12 de janeiro do próximo ano, ele não pode ser replicado. O que poderemos eventualmente considerar é em quantos dias 12 de janeiro de anos anteriores choveu e calcular uma freqüência relativa. Se tivermos essa informação, ela evidentemente poderá ser usada. Mas suponha que uma pessoa morando em Fortaleza tenha de calcular essa probabilidade. Se ela não tiver informação sobre o tempo em São Paulo, poderá simplesmente dizer que essa probabilidade é de 1/2. Por outro lado, uma pessoa vivendo em São Paulo terá informações adicionais. Por exemplo, saberá que normalmente janeiro, fevereiro e março são meses com muita chuva. Esse morador de São Paulo poderá arriscar uma probabilidade, digamos de 2/3 para o evento A. Vemos, portanto, que a associação de probabilidades a um evento depende de cada indivíduo, de sua informação a respeito desse evento. Esse tipo de apreciação é particularmente recomendável quando o indivíduo julga que as replicações anteriores não sejam comparáveis com a próxima. Por exemplo, o fenômeno El Niño pode ter ocorrido com grande intensidade em janeiro de 1999, provocando muita chuva no sudeste do Brasil, e sua intensidade nos anos seguintes talvez seja menor.

Respostas a questões como essa envolvem o que chamamos de *probabilidade subjetiva*. Ou seja, cada indivíduo, baseado em informações anteriores e na sua opinião pessoal a respeito do evento em questão, pode ter uma resposta para a probabilidade desse evento. A Inferência Bayesiana, de que trataremos brevemente neste livro (veja o Capítulo 11), toma como uma de suas bases o fato de que todas as probabilidades são subjetivas. O Teorema de Bayes tem papel importante nesse tipo de inferência, pois passa a ser visto como um mecanismo de atualização de opiniões. Ou seja, o indivíduo aprende B e passa a ter opinião $P(A|B)$ sobre A.

Um ingrediente básico quando se associam probabilidades é a *coerência*. Se um indivíduo julgar que um evento A é mais provável que seu complementar, então ele deverá, como que apostando na ocorrência de A, associar uma probabilidade maior do que 1/2 ao evento A. Por exemplo, se ele julgar que uma proporção 3 : 1 a favor de A é razoável, então ele deverá sugerir $P(A) = 3/4$. A fórmula de Bayes fornece uma maneira coerente de atualizar opiniões.

As probabilidades associadas a eventos de modo subjetivo têm propriedades análogas àquelas vistas em seções anteriores e podem ser obtidas a partir do princípio da coerência. Há outras maneiras de se associar probabilidades a eventos e os interessados poderão consultar O'Hagan (1994), por exemplo, para obter mais informações sobre esse assunto e outros ligados à Inferência Bayesiana.

5.6 Problemas e Complementos

26. Um restaurante popular apresenta apenas dois tipos de refeições: salada completa ou um prato à base de carne. Considere que 20% dos fregueses do sexo masculino preferem a salada, 30% das mulheres escolhem carne, 75% dos fregueses são homens e os seguintes eventos:

    H: freguês é homem — A: freguês prefere salada

    M: freguês é mulher — B: freguês prefere carne

    Calcular:

    (a) $P(H), P(A|H), P(B|M)$; (b) $P(A \cap H), P(A \cup H)$; (c) $P(M|A)$.

27. Uma companhia de seguros analisou a freqüência com que 2.000 segurados (1.000 homens e 1.000 mulheres) usaram o hospital. Os resultados são apresentados na tabela:

	Homens	Mulheres
Usaram o hospital	100	150
Não usaram o hospital	900	850

    (a) Qual a probabilidade de que uma pessoa segurada use o hospital?
    (b) O uso do hospital independe do sexo do segurado?

28. As probabilidades de três motoristas serem capazes de guiar até em casa com segurança, depois de beber, são de 1/3, 1/4 e 1/5, respectivamente. Se decidirem guiar até em casa, depois de beber numa festa, qual a probabilidade de todos os três motoristas sofrerem acidentes? Qual a probabilidade de pelo menos um dos motoristas guiar até em casa a salvo?

29. Duas lâmpadas queimadas foram acidentalmente misturadas com seis lâmpadas boas. Se vamos testando as lâmpadas, uma por uma, até encontrar duas defeituosas, qual é a probabilidade de que a última defeituosa seja encontrada no quarto teste?

30. Suponhamos que 10.000 bilhetes sejam vendidos em uma loteria e 5.000 em outra, cada uma tendo apenas um ganhador. Um homem tem 100 bilhetes de cada. Qual a probabilidade de que:

    (a) ele ganhe exatamente um prêmio?
    (b) ele ganhe alguma coisa?

Capítulo 6

Variáveis Aleatórias Discretas

6.1 Introdução

No capítulo anterior introduzimos alguns modelos probabilísticos por meio de espaços amostrais bem simples. Isso facilitou bastante a compreensão do conceito de probabilidade e a obtenção de algumas propriedades. Mas, para atender a situações práticas mais gerais, necessitamos ampliar esses conceitos para que tenhamos modelos probabilísticos que representem todos os tipos de variáveis definidas no Capítulo 2. Muito do que foi apresentado naquele capítulo para tratamento descritivo das variáveis terá o seu correspondente no modelo teórico.

Para as variáveis qualitativas, a descrição de probabilidades associadas a eventos construída no capítulo precedente adapta-se muito bem. Dada a sua simplicidade, trataremos aqui de variáveis quantitativas discretas. Já os modelos para variáveis contínuas necessitarão de um artifício matemático, baseado em uma generalização do conceito de histograma, definido na seção 2.3, e esse será o objetivo do próximo capítulo. A extensão dos modelos para várias variáveis será tratada no Capítulo 8.

Por outro lado, quando estudamos a descrição de dados, vimos que os recursos disponíveis para a análise das variáveis quantitativas são muito mais ricos do que para as variáveis qualitativas. Isso sugere o uso de artifícios para transformar essas últimas variáveis naquelas do primeiro tipo. Por exemplo, considere o caso de um questionário em que uma pessoa é indagada a respeito de uma proposição, e as respostas possíveis são *sim* ou *não*. Podemos associar ao problema uma variável que toma dois valores, 1 ou 0, por exemplo, correspondentes às respostas *sim* ou *não*, respectivamente. Esse tipo de variável será estudado neste capítulo.

O conhecimento de modelos probabilísticos para variáveis quantitativas é muito importante, e grande parte do restante deste livro será dedicada à construção desses modelos e inferências sobre seus parâmetros. Essas variáveis, para as quais iremos construir modelos probabilísticos, serão chamadas de *variáveis aleatórias* (v.a.).

6.2 O Conceito de Variável Aleatória Discreta

O conceito de v.a. discreta será introduzido por meio de um exemplo.

Exemplo 6.1. Um empresário pretende estabelecer uma firma para montagem de um produto composto de uma esfera e um cilindro. As partes são adquiridas em fábricas diferentes (A e B), e a montagem consistirá em juntar as duas partes e pintá-las. O produto acabado deve ter o comprimento (definido pelo cilindro) e a espessura (definida pela esfera) dentro de certos limites, e isso só poderá ser verificado após a montagem. Para estudar a viabilidade de seu empreendimento, o empresário quer ter uma idéia da distribuição do lucro por peça montada.

Sabe-se que cada componente pode ser classificado como bom, longo ou curto, conforme sua medida esteja dentro da especificação, maior ou menor que a especificada, respectivamente. Além disso, foram obtidos dos fabricantes o preço de cada componente ($5,00) e as probabilidades de produção de cada componente com as características bom, longo e curto. Esses valores estão na Tabela 6.1.

Se o produto final apresentar algum componente com a característica C (curto), ele será irrecuperável, e o conjunto será vendido como sucata ao preço de $5,00. Cada componente longo poderá ser recuperado a um custo adicional de $5,00. Se o preço de venda de cada unidade for de $25,00, como seria a distribuição de freqüências da variável X: lucro por conjunto montado?

Tabela 6.1: Distribuição da produção das fábricas A e B, de acordo com as medidas das peças produzidas.

Produto		Fábrica A Cilindro	Fábrica B Esfera
Dentro das especificações	bom (B)	0,80	0,70
Maior que as especificações	longo (L)	0,10	0,20
Menor que as especificações	curto (C)	0,10	0,10

Fonte: Retirada das especificações técnicas das fábricas *A* e *B*.

A construção dessa distribuição de freqüências vai depender de certas *suposições* que faremos sobre o comportamento do sistema considerado. Com base nessas suposições, estaremos trabalhando com um *modelo* da realidade, e a distribuição que obtivermos será uma distribuição teórica, tanto mais próxima da distribuição de freqüências real quanto mais fiéis à realidade forem as suposições.

Primeiramente, vejamos a construção do espaço amostral para a montagem dos conjuntos segundo as características de cada componente e suas respectivas probabilidades. Como os componentes vêm de fábricas diferentes, vamos supor que a classificação dos cilindros e a da esfera, segundo suas características, sejam eventos independentes. Obteremos a configuração da Figura 6.1.

Uma representação do espaço amostral em questão está apresentada na Tabela 6.2 e foi obtida da Figura 6.1.

Figura 6.1: Diagrama em árvore para o Exemplo 6.1.

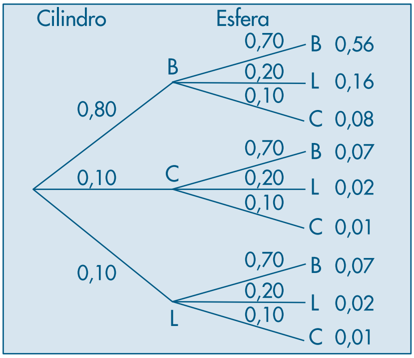


Tabela 6.2: Distribuição de probabilidade das possíveis composições das montagens.

Produto	Probabilidade	Lucro por montagem (X)
BB	0,56	15
BL	0,16	10
BC	0,08	–5
LB	0,07	10
LL	0,02	5
LC	0,01	–5
CB	0,07	–5
CL	0,02	–5
CC	0,01	–5

Fonte: Figura 5.1 e informações no texto.

A última coluna da Tabela 6.2 foi construída com base nas informações sobre preços. Por exemplo, obtendo uma montagem LB (cilindro longo e esfera boa), do preço de venda \$25,00 devemos descontar: \$10,00 dos custos dos componentes e \$5,00 para recuperar o cilindro longo. Portanto, o lucro X desse conjunto será \$10,00. Verifique os lucros das demais montagens.

Com os dados da Tabela 6.2, vemos que X pode assumir um dos seguintes valores:

$$\begin{aligned} 15&, \text{ se ocorrer o evento } A_1 = \{BB\};\\ 10&, \text{ se ocorrer o evento } A_2 = \{BL, LB\};\\ 5&, \text{ se ocorrer o evento } A_3 = \{LL\};\\ -5&, \text{ se ocorrer o evento } A_4 = \{BC, LC, CB, CL, CC\}. \end{aligned}$$

Cada um desses eventos tem uma probabilidade associada, ou seja,

$$P(A_1) = 0{,}56, \quad P(A_2) = 0{,}23,$$
$$P(A_3) = 0{,}02, \quad P(A_4) = 0{,}19,$$

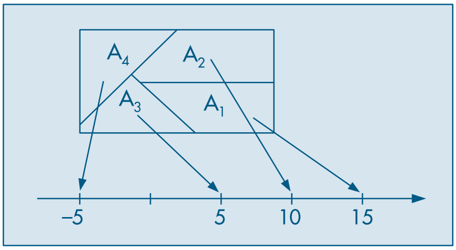
o que nos permite escrever a função $(x, p(x))$ da Tabela 6.3, que é um modelo teórico para a distribuição da variável X, que o empresário poderá usar para julgar a viabilidade econômica do projeto que ele pretende realizar. Aqui, x é o valor da v.a. X e $p(x)$ é a probabilidade de X tomar o valor x. Voltaremos a esse problema mais adiante.

Tabela 6.3: Distribuição da v.a. X.

x	$p(x)$
15	0,56
10	0,23
5	0,02
–5	0,19
Total	1,00

A função $(x, p(x))$ é chamada *função de probabilidade* da v.a. X. Esquematicamente teremos a situação da Figura 6.2.

Figura 6.2: Função de probabilidade da v.a. X = lucro por montagem.



É evidente que, ao mesmo espaço amostral da Tabela 6.2, podemos associar outras variáveis aleatórias, como veremos a seguir.

Exemplo 6.2. Se considerarmos Y como sendo a variável "custo de recuperação de cada conjunto produzido", verificaremos que Y irá assumir os valores

$$\begin{aligned} &0, \text{ se ocorrer o evento } B_1 = \{BB, BC, LC, CB, CL, CC\};\\ &5, \text{ se ocorrer o evento } B_2 = \{BL, LB\};\\ &10, \text{ se ocorrer o evento } B_3 = \{LL\}. \end{aligned}$$

A função de probabilidade da v.a. Y está representada na Tabela 6.4 e a Figura 6.3 representa a situação esquematicamente.

Figura 6.3: Função de probabilidade da v.a. Y = custo de recuperação.

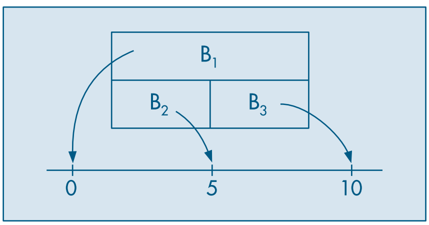


Tabela 6.4: Distribuição da v.a. Y.

y	$p(y)$
0	0,75
5	0,23
10	0,02
Total	1,00

Deduz-se do exposto que uma v.a. X, do tipo discreto, estará bem caracterizada se indicarmos os possíveis valores x_1, x_2, ..., x_n, ... que ela pode assumir e as respectivas probabilidades $p(x_1)$, $p(x_2)$, ..., $p(x_n)$, ..., ou seja, se conhecermos a sua função de probabilidade $(x, p(x))$. Também usaremos a notação $p(x) = P(X = x)$.

Em algumas situações, a determinação da função de probabilidade (f.p.) é bem mais simples. Isso pode ser verificado pelos dois exemplos seguintes.

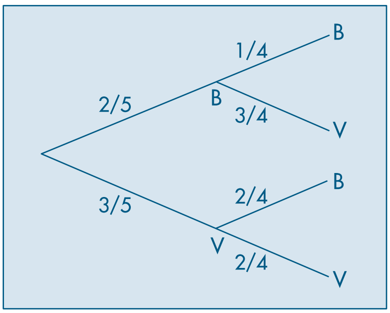
Exemplo 6.3. Voltemos à situação do Exemplo 5.10, em que consideramos duas extrações, sem reposição, de uma urna contendo duas bolas brancas e três bolas vermelhas. Definamos a v.a. X: número de bolas vermelhas obtidas nas duas extrações. Obtemos a Tabela 6.5 e a Figura 6.4.

Tabela 6.5: Extrações sem reposição de urna com duas bolas brancas e três bolas vermelhas.

Resultados	Probabilidades	X
BB	1/10	0
BV	3/10	1
VB	3/10	1
VV	3/10	2

Fonte: Figura 6.4.

Figura 6.4: Diagrama em árvore para o Exemplo 6.3.



Vemos, pois, que a cada resultado do experimento está associado um valor da v.a. X, a saber, 0, 1 ou 2.

Temos que $X = 0$, com probabilidade 1/10, pois $X = 0$ se, e somente se, ocorre o resultado BB; $X = 1$ com probabilidade 3/10 + 3/10 = 6/10, pois $X = 1$ se, e somente se, ocorrem os resultados BV ou VB, que são mutuamente exclusivos; finalmente, $X = 2$ com probabilidade 3/10, pois $X = 2$ se, e somente se, ocorre o resultado VV. Resumidamente,

$$p(0) = P(X = 0) = P(BB) = 1/10,$$
$$p(1) = P(X = 1) = P(BV \text{ ou } VB) = 6/10,$$
$$p(2) = P(X = 2) = P(VV) = 3/10.$$

Na Tabela 6.6 apresentamos a distribuição de probabilidades da v.a. X.

Tabela 6.6: Distribuição de probabilidades da v.a. X = número de bolas vermelhas.

x	$p(x)$
0	1/10
1	6/10
2	3/10

Fonte: Tabela 6.5.

Exemplo 6.4. Retomemos o Exemplo 5.3, em que consideramos o lançamento de uma moeda duas vezes. Definamos a v.a. Y: número de caras obtidas nos dois lançamentos. Temos, então:

$$p(0) = P(Y = 0) = P(RR) = 1/4,$$
$$p(1) = P(Y = 1) = P(CR \text{ ou } RC) = 1/4 + 1/4 = 1/2,$$
$$p(2) = P(Y = 2) = P(CC) = 1/4.$$

Na Tabela 6.7 e Figura 6.5 temos esquematizado o que ocorre e na Tabela 6.8 apresentamos a distribuição de probabilidades de Y.

Tabela 6.7: Lançamento de duas moedas.

Resultados	Probabilidades	Y
CC	1/4	2
CR	1/4	1
RC	1/4	1
RR	1/4	0

Fonte: Figura 6.5.

Figura 6.5: Diagrama em árvore para o Exemplo 6.4.

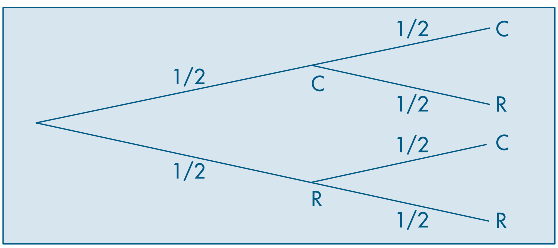


Tabela 6.8: Distribuição da v.a. Y = número de caras.

y	$p(y)$
0	1/4
1	1/2
2	1/4

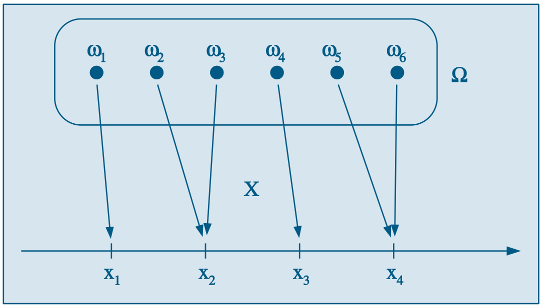
Fonte: Tabela 6.7.

Dos exemplos apresentados, vemos que, a cada ponto do espaço amostral, a variável sob consideração associa um valor numérico, o que corresponde em Matemática ao conceito de função, mais precisamente, a uma função definida no espaço amostral Ω e assumindo valores reais.

Definição. Uma função X, definida no espaço amostral Ω e com valores num conjunto enumerável de pontos da reta é dita uma variável aleatória discreta.

Esquematicamente, teremos a situação da Figura 6.6.

Figura 6.6: Definição de uma v.a.



Vimos, também, como associar a cada valor x_i da v.a. X sua probabilidade de ocorrência. Ela é dada pela probabilidade do evento A de Ω, cujos elementos correspondem ao valor x_i (veja Figuras 6.2 e 6.3). Matematicamente, podemos escrever

$$P(X = x_i) = P(A),$$

onde

$$A = \{\omega_1, \omega_2, ...\} \subset \Omega$$

é tal que $X(\omega_i) = x_i$, se $\omega_i \in A$ e $X(\omega_i) \neq x_i$, se $\omega_i \in A^c$.

Definição. Chama-se função de probabilidade da v.a. discreta X, que assume os valores $x_1, x_2, ..., x_n, ...$, a função $\{(x_i, p(x_i)), i = 1, 2, ...\}$, que a cada valor de x_i associa a sua probabilidade de ocorrência, isto é,

$$p(x_i) = P(X = x_i) = p_i, \ i = 1, 2, ...$$

Problemas

1. Considere uma urna contendo três bolas vermelhas e cinco pretas. Retire três bolas, sem reposição, e defina a v.a. X igual ao número de bolas pretas. Obtenha a distribuição de X.

2. Repita o problema anterior, mas considerando extrações com reposição.

3. Suponha que uma moeda perfeita é lançada até que cara apareça pela primeira vez. Seja X o número de lançamentos até que isso aconteça. Obtenha a distribuição de X. (Observe que, nesse problema, pelo menos teoricamente, X pode assumir um número infinito de valores.) Veja também o Problema 55.

4. Uma moeda perfeita é lançada quatro vezes. Seja Y o número de caras obtidas. Calcule a distribuição de Y.

5. Repita o problema anterior, considerando agora que a moeda é viciada, sendo a probabilidade de cara dada por p, $0 < p < 1$, $p \neq 1/2$.

6. Generalize o Problema 5, para n lançamentos da moeda.

6.3 Valor Médio de uma Variável Aleatória

Vamos introduzir o conceito de valor médio por meio do seguinte exemplo.

Exemplo 6.5. Uma pergunta que logo ocorreria ao empresário do Exemplo 6.1 é qual o lucro médio por conjunto montado que ele espera conseguir. Da Tabela 6.3, observamos que 56% das montagens devem produzir um lucro de 15 reais, 23% um lucro de dez reais, e assim por diante. Logo, o lucro esperado por montagem será dado por

$$\text{lucro médio} = (0{,}56)(15) + (0{,}23)(10) + (0{,}02)(5) + (0{,}19)(-5) = 9{,}85.$$

Isto é, caso sejam verdadeiras as suposições feitas para determinar a distribuição da v.a., o empresário espera ter um lucro de 9,85 reais por conjunto montado.

Definição. Dada a v.a. X discreta, assumindo os valores $x_1, \ldots, x_n$, chamamos valor médio ou esperança matemática de X ao valor

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i) = \sum_{i=1}^{n} x_i p_i. \tag{6.1}$$

A expressão (6.1) é semelhante àquela utilizada para a média, introduzida no Capítulo 3, onde no lugar das probabilidades p_i tínhamos as freqüências relativas f_i. A distinção entre essas duas quantidades é que a primeira corresponde a valores de um modelo teórico pressuposto, e a segunda, a valores observados da variável. Como p_i e f_i têm a mesma interpretação, todas as medidas e gráficos discutidos no Capítulo 2, baseados na distribuição das f_i, possuem um correspondente na distribuição de uma v.a. Além do valor médio, ou simplesmente *média*, definido acima, podemos considerar também outras medidas de posição e variabilidade, como a mediana e o desvio padrão. Veja a seção 6.8 para a definição da mediana de uma v.a. discreta. Vamos considerar agora a definição de variância.

Definição. Chamamos de *variância* da v.a. X o valor

$$\mathrm{Var}(X) = \sum_{i=1}^{n} [x_i - E(X)]^2 p_i. \tag{6.2}$$

O desvio padrão de X, DP(X), é definido como a raiz quadrada positiva da variância.

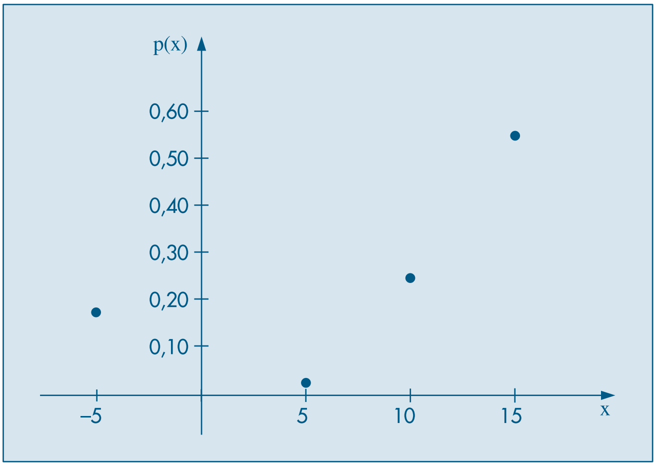
Exemplo 6.6. Deixamos a cargo do leitor verificar que, no caso do problema do empresário, teremos:

(i) Var(X) = 57,23;

(ii) $DP(X)$ = 7,57;

(iii) gráfico de (x, $p(x)$): Figura 6.7.

Figura 6.7: Gráfico de $p(x)$: distribuição da v.a. X = lucro por montagem.



Observação. Até agora, consideramos o caso em que a v.a. X pode assumir um número *finito* de valores. Mas uma v.a. discreta X pode assumir um número *infinito*, porém *enumerável*, de valores, $x_1, \ldots, x_n, \ldots$, com probabilidades $p_1, \ldots, p_n, \ldots$, tal que cada $p_i > 0$ e a soma de todos os p_i seja 1, ou seja, $\sum_{i=1}^{\infty} p_i = 1$. Veja o Problema 3. Nesse caso, a definição de esperança deve ser modificada. A soma na expressão (6.1) é uma "soma infinita", que temos de supor que seja "convergente".

Problemas

7. Obtenha a média e a variância da v.a. X dos Problemas 1 e 2.

8. Obter a média e a variância da v.a. Y do Problema 4.

6.4 Algumas Propriedades do Valor Médio

Retomemos o Exemplo 6.1 para ilustrar algumas propriedades da média de uma v.a.

Exemplo 6.7. Suponha que todos os preços determinados pelo empresário do Exemplo 6.1 estivessem errados. Na realidade, todos os valores deveriam ser duplicados, isto é, custos e preços de venda. Isso corresponde à transformação $Z = 2X$. As probabilidades associadas à v.a. Z serão as mesmas da v.a. X, pois cada valor de X irá corresponder a um único valor de Z. Na Tabela 6.9 temos a distribuição de Z.

O valor médio da v.a. Z é obtido por

$$E(Z) = \sum z_i p(z_i) = \sum (2x_i) p(x_i) = 19{,}70.$$

Suponha, agora, que queiramos a distribuição da v.a. $W = X^2$. Baseados na Tabela 6.3, obtemos a Tabela 6.10.

Tabela 6.9: Distribuição da variável aleatória $Z = 2X$.

x	$z = 2x$	$p(z) = p(x)$	$z \cdot p(z)$
15	30	0,56	16,80
10	20	0,23	4,60
5	10	0,02	0,20
–5	–10	0,19	–1,90
Total	—	1,00	19,70

Fonte: Tabela 6.3.

Tabela 6.10: Distribuição da variável aleatória $W = X^2$.

w	$p(w)$	$w \cdot p(w)$
225	0,56	126,00
100	0,23	23,00
25	0,21	5,25
Total	1,00	154,25

Fonte: Tabela 6.3.

Observe que o evento $\{W = 25\}$ ocorre quando $\{X = 5 \text{ ou } X = -5\}$, portanto $P(W = 25) = P(X = 5) + P(X = -5) = 0{,}02 + 0{,}19 = 0{,}21$. Segue-se que a média de W é

$$\begin{aligned} E(W) &= \sum w_i p(w_i) = (225)(0{,}56) + (100)(0{,}23) + (25)(0{,}21) \\ &= (225)(0{,}56) + (100)(0{,}23) + \{(25)(0{,}02) + (25)(0{,}19)\} \\ &= \sum x_i^2 p(x_i) = 154{,}25. \end{aligned}$$

Quanto às esperanças de Z e W, transformadas de X, é fácil ver que elas podem ser escritas através da f.p. de X.

Definição. Dada a v.a. discreta X e a respectiva função de probabilidade $p(x)$, a esperança matemática da função $h(X)$ é dada por

$$E[h(X)] = \sum h(x_i)p(x_i). \tag{6.3}$$

As seguintes propriedades podem ser facilmente demonstradas (veja o Problema 45):

(a) Se $h(X) = aX + b$, onde a e b são constantes, então

$$E(aX + b) = aE(X) + b, \tag{6.4}$$

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X). \tag{6.5}$$

(b) $\mathrm{Var}(X) = E(X^2) - [(E(X)]^2 = \sum x_i^2 p(x_i) - [\sum x_i p(x_i)]^2.$ (6.6)

A fórmula (6.6) deve ser usada para facilitar o cálculo da variância.

Observação. A propriedade (6.4) não vale, em geral, para funções não-lineares. Veja o Problema 58.

Exemplo 6.8. Usando os resultados dos exemplos 6.5 e 6.7, obtemos

$$\mathrm{Var}(X) = 154{,}25 - (9{,}85)^2 = 57{,}23.$$

Observação. Usaremos os símbolos abaixo para indicar a média e a variância de uma v.a. X:

$$E(X) = \mu(X),$$
$$\mathrm{Var}(X) = \sigma^2(X),$$

ou, simplesmente, μ e σ^2, respectivamente, se não houver possibilidade de confusão.

6.5 Função de Distribuição Acumulada

No Capítulo 2 demos a definição de função de distribuição acumulada ou empírica para um conjunto de n observações. O equivalente teórico para variáveis aleatórias é definido a seguir.

Definição. Dada a variável aleatória X, chamaremos de função de distribuição acumulada (f.d.a.), ou simplesmente função de distribuição (f.d.) $F(x)$ à função

$$F(x) = P(X \leqslant x). \tag{6.7}$$

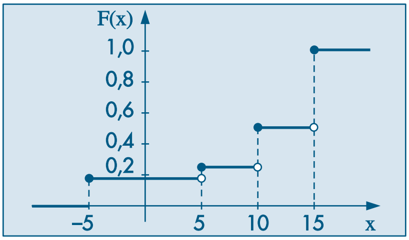
Observe que o domínio de F é todo o conjunto dos números reais, ao passo que o contradomínio é o intervalo [0,1].

Exemplo 6.9. Voltando ao problema do empresário e usando a f.p. de X definida na Tabela 6.3, a f.d.a. de X será dada por

$$F(x) = \begin{cases} 0, & \text{se } x < -5 \\ 0{,}19, & \text{se } -5 \leqslant x < 5 \\ 0{,}21, & \text{se } 5 \leqslant x < 10 \\ 0{,}44, & \text{se } 10 \leqslant x < 15 \\ 1, & \text{se } x \geqslant 15, \end{cases}$$

cujo gráfico está na Figura 6.8.

Figura 6.8: f.d.a. para a v.a. X = lucro por montagem.



Observe que $P(X = x_i)$ é igual ao salto que a função $F(x)$ dá no ponto x_i; por exemplo, $P(X = 10) = 0{,}23 = F(10) - F(10-)$. De modo geral, $P(X = x_i) = F(x_i) - F(x_i-)$, onde lembramos que $F(a-) = \lim_{x \to a\_} F(x)$. Observe, também, que o conhecimento de $F(x)$ é equivalente ao conhecimento da f.p. de X.

Problemas

9. No Problema 1, obtenha as distribuições das v.a. $3X$ e X^2.

10. Considere o lançamento de três moedas. Se ocorre o evento CCC, dizemos que temos uma *seqüência*, ao passo que se ocorre o evento CRC temos *três seqüências*. Defina a v.a. X = número de caras obtidas e Y = número de seqüências, isso para cada resultado possível. Assim, $X(CRR) = 1$ e $Y(CRR) = 2$. Obtenha as distribuições de X e Y. Calcule $E(X)$, $E(Y)$, $\text{Var}(X)$ e $\text{Var}(Y)$.

11. Suponha que a v.a. V tem a distribuição seguinte:

v	0	1
$p(v)$	q	$1-q$

Obtenha $E(V)$ e $\text{Var}(V)$.

12. Seja X com distribuição dada abaixo; calcule $E(X)$. Considere a v.a. $(X-a)^2$ e calcule $E(X-a)^2$ para $a = 0$, 1/4, 1/2, 3/4, 1. Obtenha o gráfico de $E(X-a)^2 = g(a)$. Para qual valor de a, $g(a)$ é mínimo?

x	0	1	2
$p(x)$	1/2	1/4	1/4

13. Um vendedor de equipamento pesado pode visitar, num dia, um ou dois clientes, com probabilidade de 1/3 ou 2/3, respectivamente. De cada contato, pode resultar a venda de um equipamento por \$50.000,00 (com probabilidade 1/10) ou nenhuma venda (com probabilidade 9/10). Indicando por Y o valor total de vendas diárias desse vendedor, escreva a função de probabilidade de Y e calcule o valor total esperado de vendas diárias.

14. Calcule a variância da v.a. Y definida no Problema 13.

15. Obter a *f.d.a.* para a v.a. V do Problema 11. Faça seu gráfico.

16. Calcule a *f.d.a.* da v.a. Y do Problema 10 e faça seu gráfico.

17. O tempo T, em minutos, necessário para um operário processar certa peça é uma v.a. com a seguinte distribuição de probabilidade.

t	2	3	4	5	6	7
$p(t)$	0,1	0,1	0,3	0,2	0,2	0,1

(*a*) Calcule o tempo médio de processamento.
Para cada peça processada, o operário ganha um fixo de \$2,00, mas, se ele processa a peça em menos de seis minutos, ganha \$0,50 em cada minuto poupado. Por exemplo, se ele processa a peça em quatro minutos, recebe a quantia adicional de \$1,00.
(*b*) Encontre a distribuição, a média e a variância da v.a. G: quantia em \$ ganha por peça.

18. Sabe-se que a v.a. X assume os valores 1, 2 e 3 e que sua *f.d.a.* $F(x)$ é tal que

$$F(1) - F(1-) = 1/3,$$
$$F(2) - F(2-) = 1/6,$$
$$F(3) - F(3-) = 1/2.$$

Obtenha a distribuição de X, a *f.d.a.* $F(x)$ e os gráficos respectivos.

19. Obtenha a *f.d.a.* $F(t)$ da v.a. T do Problema 17.

6.6 Alguns Modelos Probabilísticos para Variáveis Aleatórias Discretas

Algumas variáveis aleatórias adaptam-se muito bem a uma série de problemas práticos. Portanto, um estudo pormenorizado dessas variáveis é de grande importância para a construção de modelos probabilísticos para situações reais e a conseqüente estimação de seus parâmetros. Para algumas dessas distribuições existem tabelas que facilitam o cálculo de probabilidades, em função de seus parâmetros. Nesta seção iremos estudar alguns desses modelos, procurando enfatizar as condições em que eles aparecem, suas funções de probabilidade, parâmetros e como calcular probabilidades.

6.6.1 Distribuição Uniforme Discreta

Este é o caso mais simples de v.a. discreta, em que cada valor possível ocorre com a mesma probabilidade.

Definição. A v.a. discreta X, assumindo os valores $x_1, \ldots, x_k$, tem *distribuição uniforme* se, e somente se,

$$P(X = x_i) = p(x_i) = p = \frac{1}{k}, \tag{6.8}$$

para todo $i = 1, 2, \ldots, k$.

É fácil verificar que

$$E(X) = \frac{1}{k}\sum_{i=1}^{k} x_i, \tag{6.9}$$

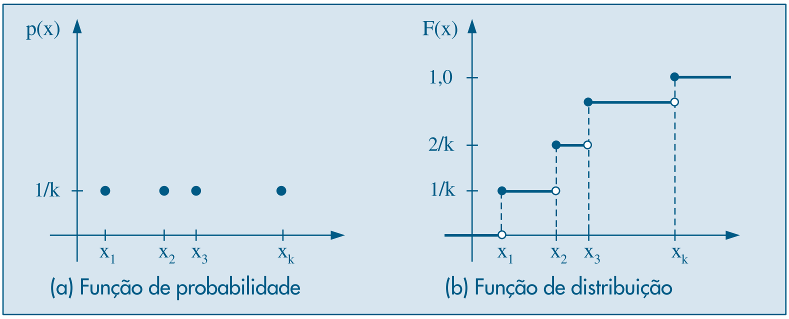
$$\mathrm{Var}(X) = \frac{1}{k}\left\{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{k}\right\}, \tag{6.10}$$

e que a função de distribuição acumulada é dada por

$$F(x) = \sum_{(x_i \leqslant x)} \frac{1}{k} = \frac{n(x)}{k}, \tag{6.11}$$

onde $n(x)$ é o número de $x_i \leqslant x$ (veja a Figura 6.9).

Figura 6.9: Distribuição uniforme discreta.



Exemplo 6.10. Seja X a v.a. que indica o "número de pontos marcados na face superior de um dado", quando ele é lançado. Obtemos na Tabela 6.11 a distribuição de X. Temos, também,

$$E(X) = 1/6\ \{1 + 2 + 3 + 4 + 5 + 6\} = 21/6 = 3{,}5,$$
$$\mathrm{Var}(X) = 1/6\ \{(1 + 4 + \ldots + 36) - (21)^2/6\} = 35/12 = 2{,}9.$$

Tabela 6.11: Número de pontos no lançamento de um dado.

x	1	2	3	4	5	6	Total
$p(x)$	1/6	1/6	1/6	1/6	1/6	1/6	1,0

6.6.2 Distribuição de Bernoulli

Muitos experimentos são tais que os resultados apresentam ou não uma determinada característica. Por exemplo:

(1) uma moeda é lançada: o resultado ou é cara, ou não (ocorrendo, então, coroa);

(2) um dado é lançado: ou ocorre face 5 ou não (ocorrendo, então, uma das faces 1, 2, 3, 4 ou 6);

(3) uma peça é escolhida ao acaso de um lote contendo 500 peças: essa peça é defeituosa ou não;

(4) uma pessoa escolhida ao acaso dentre 1.000 é ou não do sexo masculino;

(5) uma pessoa é escolhida ao acaso entre os moradores de uma cidade e verifica-se se ela é favorável ou não a um projeto municipal.

Em todos esses casos, estamos interessados na ocorrência de *sucesso* (cara, face 5 etc.) ou *fracasso* (coroa, face diferente de 5 etc.). Essa terminologia (sucesso e fracasso) será usada freqüentemente.

Para cada experimento acima, podemos definir uma v.a. X, que assume apenas dois valores: 1, se ocorrer sucesso, e 0, se ocorrer fracasso. Indicaremos por p a probabilidade de sucesso, isto é, $P(\text{sucesso}) = P(S) = p$, $0 < p < 1$.

Definição. A variável aleatória X, que assume apenas os valores 0 e 1, com função de probabilidade $(x, p(x))$ tal que

$$p(0) = P(X = 0) = 1 - p,$$
$$p(1) = P(X = 1) = p,$$

é chamada *variável aleatória de Bernoulli*.

Então, segue-se facilmente que

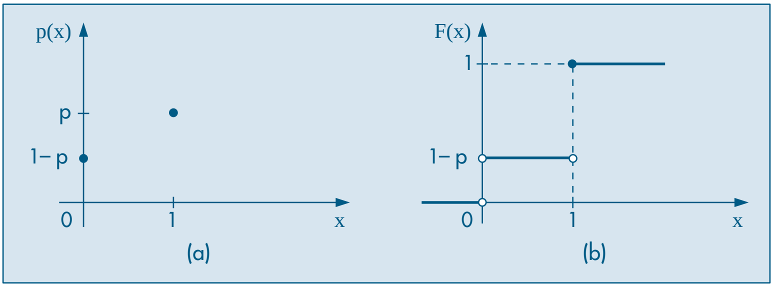
$$E(X) = p; \tag{6.12}$$

$$\text{Var}(X) = p - p^2 = p(1-p), \tag{6.13}$$

$$F(x) = \begin{cases} 0, & \text{se } x < 0 \\ 1-p, & \text{se } 0 \leqslant x < 1 \\ 1, & \text{se } x \geqslant 1. \end{cases}$$

Na Figura 6.10 temos representadas as f.p. e f.d.a. de X.

Figura 6.10: Distribuição de Bernoulli (a) f.p. (b) f.d.a.



Exemplo 6.11. Vamos supor o caso do experimento (2). Supondo o dado perfeito, teremos $P(X = 0) = 5/6$, $P(X = 1) = 1/6$,

$$E(X) = 1/6, \text{Var}(X) = (1/6)\,(5/6) = 5/36.$$

Observação. Experimentos que resultam numa v.a. de Bernoulli são chamados *ensaios de Bernoulli*. Usaremos a notação

$$X \sim \text{Ber}(p)$$

para indicar uma v.a. com distribuição de Bernoulli com parâmetro p.

6.6.3 Distribuição Binomial

Imagine, agora, que repetimos um ensaio de Bernoulli n vezes, ou, de maneira alternativa, obtemos uma amostra de tamanho n de uma distribuição de Bernoulli. Suponha ainda que as repetições sejam *independentes*, isto é, o resultado de um ensaio não tem influência nenhuma no resultado de qualquer outro ensaio. Uma amostra particular será constituída de uma seqüência de sucessos e fracassos, ou, alternativamente, de uns e zeros. Por exemplo, repetindo um ensaio de Bernoulli cinco vezes ($n = 5$), um particular resultado pode ser *FSSFS* ou a quíntupla ordenada (0, 1, 1, 0, 1). Usando a notação da seção 6.6.2, com $P(S) = p$, a probabilidade de tal amostra será

$$(1 - p)pp(1 - p)p = p^3(1 - p)^2.$$

O número de sucessos nessa amostra é igual a 3, sendo 2 o número de fracassos. Considere agora as seguintes situações, obtidas de (1) a (5) da seção anterior:

(1’) uma moeda é lançada três vezes; qual é a probabilidade de se obter duas caras?

(2’) um dado é lançado cinco vezes; qual é a probabilidade de se obter face 5 no máximo três vezes?

(3') dez peças são extraídas, ao acaso, com reposição, de um lote contendo 500 peças; qual é a probabilidade de que todas sejam defeituosas, sabendo-se que 10% das peças do lote são defeituosas?

(4') cinco pessoas são escolhidas ao acaso entre 1.000; qual é a probabilidade de que duas sejam do sexo masculino?

(5') sabe-se que 90% das pessoas de uma cidade são favoráveis a um projeto municipal. Escolhendo-se 100 pessoas ao acaso entre os moradores, qual é a probabilidade de que pelo menos 80 sejam favoráveis ao projeto?

Observe que, nos casos (4') e (5'), o fato de estarmos extraindo indivíduos de um conjunto muito grande implica que podemos supor que as extrações sejam praticamente independentes.

Exemplo 6.12. Consideremos a situação (1'), supondo que a moeda seja "honesta", isto é, $P(\text{sucesso}) = P(\text{cara}) = 1/2$. Indiquemos o sucesso (cara) por S e fracasso (coroa), por F. Então, estamos interessados na probabilidade do evento

$$A = \{SSF, SFS, FSS\},$$

ou, em termos da notação anterior, na probabilidade de

$$A = \{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}.$$

É claro que $P(A) = P(SSF) + P(SFS) + P(FSS)$ e, devido à independência dos ensaios,

$$P(SSF) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = P(SFS) = P(FSS),$$

e, portanto,

$$P(A) = \frac{3}{8}.$$

Se a probabilidade de sucesso for p, $0 < p < 1$, e $P(F) = 1 - p = q$, então

$$P(SSF) = p \times p \times q = p^2 \times q = P(SFS) = P(FSS),$$

de modo que

$$P(A) = 3p^2q.$$

Uma característica interessante dos experimentos considerados é que estamos interessados apenas no *número total* de sucessos e não na ordem em que eles ocorrem. Podemos construir a Tabela 6.12 para $n = 3$ lançamentos da moeda, com $P(S) = p$, $P(F) = 1 - p = q$, a partir da Figura 6.11.

Figura 6.11: Probabilidades binomiais para $n = 3$ e $P(S) = p$.

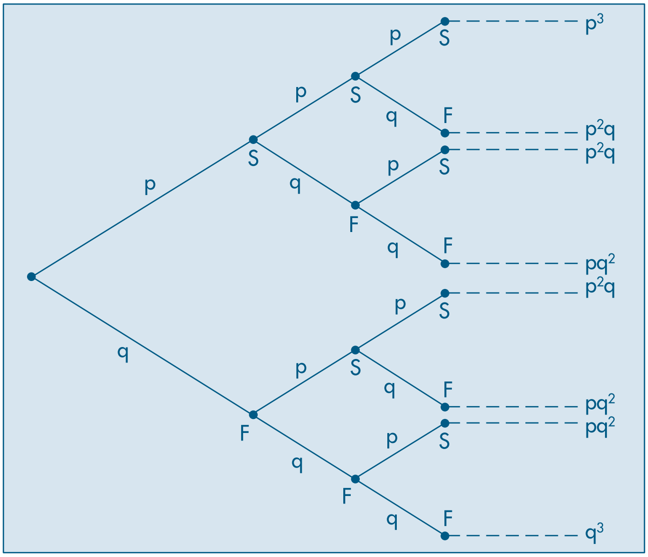


Tabela 6.12: Probabilidades binomiais para $n = 3$ e $P(S) = p$.

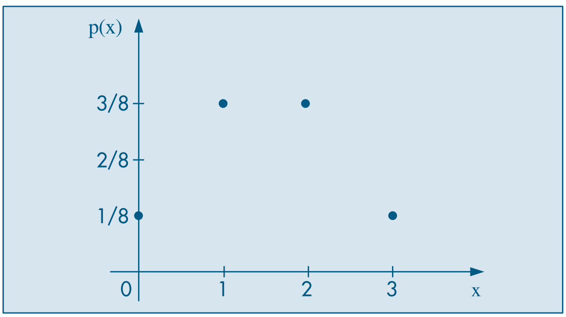
Número de sucessos	Probabilidades	$p = 1/2$
0	q^3	1/8
1	$3pq^2$	3/8
2	$3p^2q$	3/8
3	p^3	1/8

Fonte: Figura 6.11.

Vamos designar por X o número total de sucessos em n ensaios de Bernoulli, com probabilidade de sucesso p, $0 < p < 1$. Os possíveis valores de X são 0, 1, 2, ..., n e os pares $(x, p(x))$, onde $p(x) = P(X = x)$, constituem a chamada *distribuição binomial*.

Para o exemplo (1') acima, $n = 3$ e $p = 1/2$, obtemos a distribuição dada pela primeira e terceira colunas da Tabela 6.12 e o gráfico da Figura 6.12.

Figura 6.12: Gráfico da f.p. $p(x)$ para $n = 3$ e $p = 1/2$.



Obtenhamos, agora, $P(X = k)$, ou seja, numa seqüência de n ensaios de Bernoulli, a probabilidade de obter k sucessos (e portanto $n - k$ fracassos), $k = 0,1,2, \ldots, n$, com $P(S) = p$, $P(F) = 1 - p = q$. Uma particular seqüência é

$$SSS \ldots SFF \ldots F,$$

onde temos k sucessos seguidos por $n - k$ fracassos. A probabilidade de tal seqüência é

$$p^k(1-p)^{n-k} = p^k q^{n-k}, \tag{6.14}$$

devido à independência dos ensaios. Mas *qualquer* seqüência com k sucessos e $n - k$ fracassos terá a mesma probabilidade (6.14). Portanto resta saber quantas seqüências com a propriedade especificada podemos formar. É fácil ver que existem

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

tais seqüências, de modo que

$$P(X = k) = \binom{n}{k} p^k q^{n-k},\ k = 0, 1, \ldots, n. \tag{6.15}$$

As probabilidades (6.15) também serão indicadas por $b(k; n, p)$ e, quando a v.a. X tiver distribuição binomial com parâmetros n e p, escreveremos

$$X \sim b(n, p).$$

Exemplo 6.13. Vamos considerar a situação (3′) acima. Temos $n = 10$ ensaios de Bernoulli, cada um com $P(S) = P(\text{peça defeituosa}) = p = 0{,}1$. Se X indicar o número de peças defeituosas na amostra, queremos calcular $P(X = 10) = b(10; 10, 1/10)$. Por (6.15), obtemos

$$P(X = 10) = \binom{10}{10}(1/10)^{10}(9/10)^0 = (1/10)^{10} = 1/10^{10}.$$

A média e a variância de uma v.a. binomial, com parâmetros n e p são dadas, respectivamente, por

$$E(X) = np, \tag{6.16}$$

$$\operatorname{Var}(X) = npq. \tag{6.17}$$

Veja o Problema 41 e as seções 8.3 e 8.4.

Para o Exemplo 6.13 temos

$$E(X) = 10 \times \frac{1}{10} = 1,$$

$$\operatorname{Var}(X) = 10 \times \frac{1}{10} \times \frac{9}{10} = \frac{9}{10}.$$

As probabilidades binomiais $b(k; n, p)$ são facilmente calculadas em programas estatísticos, como o Minitab e o SPlus, ou planilhas, como o Excel, ou então são dadas por tabelas especialmente construídas, para diferentes valores de n e p. A Tabela I fornece essas probabilidades para valores de $n = 2, 3, ..., 19$ e alguns valores de p.

Exemplo 6.14. Usando (6.15) e a Tabela I, ou com a ajuda de um computador, obtemos

$$b(17;\ 20;\ 0{,}9) = \binom{20}{17}(0{,}9)^{17}\ (0{,}1)^{3} = 0{,}19.$$

No Capítulo 7 e seção 6.6.5 abaixo veremos duas maneiras de calcular valores aproximados para as probabilidades binomiais para n grande.

Para finalizar, vamos formalizar os principais pontos apresentados nesta seção.

Definição. Chama-se de *experimento binomial* ao experimento
(a) que consiste em n ensaios de Bernoulli;
(b) cujos ensaios são independentes; e
(c) para o qual a probabilidade de sucesso em cada ensaio é sempre igual a p, $0 < p < 1$.

Definição. A variável aleatória X, correspondente ao número de sucessos num experimento binomial, tem distribuição binomial $b(n, p)$, com função de probabilidade

$$b(k;\ n,\ p) = P(X = k|n,\ p) = \binom{n}{k} p^k q^{n-k},\quad k = 0,\ 1,\ ...,\ n. \tag{6.18}$$

Na seção 6.9 veremos como podemos obter os valores $b(k;\ n,\ p)$, para n e p dados, usando um pacote estatístico.

6.6.4 Distribuição Hipergeométrica

Essa distribuição é adequada quando consideramos extrações casuais feitas *sem reposição* de uma população dividida segundo dois atributos. Para ilustrar, considere uma população de N objetos, r dos quais têm o atributo A e $N - r$ têm o atributo B. Um grupo de n elementos é escolhido ao acaso, sem reposição. Estamos interessados em calcular a probabilidade de que esse grupo contenha k elementos com o atributo A. Pode-se ver facilmente, utilizando o princípio multiplicativo, que essa probabilidade é dada por

$$p_k = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}, \tag{6.19}$$

onde $\max(0,\ n - N + r) \leqslant k \leqslant \min(r,\ n)$.

Os pares $(k,\ p_k)$ constituem a *distribuição hipergeométrica de probabilidades*. Se definirmos a v.a. X como sendo o número de elementos na amostra que têm o atributo A, então $P(X = k) = p_k$.

Exemplo 6.15. Em problemas de controle de qualidade, lotes com N itens são examinados. O número de itens com defeito (atributo A), r, é desconhecido. Colhemos uma amostra de n

itens e determinamos k. Somente para ilustrar, suponha que num lote de $N = 100$ peças, $r = 10$ sejam defeituosas. Escolhendo $n = 5$ peças sem reposição, a probabilidade de não se obter peças defeituosas é

$$p_0 = \frac{\binom{10}{0}\binom{90}{5}}{\binom{100}{5}} = \frac{\binom{90}{5}}{\binom{100}{5}} \simeq 0{,}584,$$

enquanto a probabilidade de se obter pelo menos uma defeituosa é

$$p_1 + p_2 + \dots + p_5 = 1 - p_0 \simeq 0{,}426.$$

Pode-se demonstrar que a v.a. X definida acima tem esperança e variância dadas por

$$E(X) = np, \tag{6.20}$$

$$\text{Var}(X) = np(1-p)\,\frac{N-n}{N-1}, \tag{6.21}$$

respectivamente, onde $p = r/N$ é a probabilidade de se obter uma peça defeituosa numa única extração. Se N for grande, quando comparado com n, então extrações com ou sem reposição serão praticamente equivalentes, de modo que as probabilidades dadas por (6.19) serão aproximadamente iguais às dadas pela fórmula (6.15), isto é, $p_k \simeq b(k;\, n,\, p)$. Do mesmo modo, os resultados (6.20) e (6.21) serão aproximadamente iguais aos valores correspondentes da distribuição binomial (note que $N - n \simeq N - 1$, se $n << N$). Denotaremos uma v.a. com distribuição hipergeométrica por

$$X \sim \text{hip}(N,\, r,\, n).$$

6.6.5 Distribuição de Poisson

A Tabela I fornece os valores de $b(k;\, n,\, p)$ para $n = 2, \dots, 19$. Para n grande e p pequeno, podemos aproximar essas probabilidades por

$$\frac{e^{-np}(np)^k}{k!},\quad k = 0,\, 1,\, \dots,\, n. \tag{6.22}$$

As probabilidades (6.22), calculadas agora para todos os valores inteiros não negativos $k = 0, 1, 2, \dots$, constituem a chamada *distribuição de Poisson*, tabelada na Tabela II, para alguns valores de $\lambda = np$. A aproximação

$$b(k;\, n,\, p) \simeq \frac{e^{-np}(np)^k}{k!} \tag{6.23}$$

é boa se n for grande e p pequeno e de tal sorte que $np \leqslant 7$. Ver o Problema 43 para uma sugestão de como provar (6.23).

As probabilidades dadas por (6.23) podem, também, ser obtidas em aplicativos estatísticos ou planilhas, assim como a binomial.

Exemplo 6.16. Consideremos aproximar $b(2;\ 1.000,\ 0{,}0001)$, usando (6.23). Temos que $np = 0{,}1$, logo

$$b(2;\ 1.000,\ 0{,}0001) \simeq \frac{e^{-0,1}(0{,}1)^2}{2!} = 0{,}0045.$$

Observemos que as probabilidades (6.23) estão definidas para qualquer inteiro não negativo k. Contudo, observando a Tabela II, vemos que essas probabilidades decaem à medida que k cresce e, normalmente, são desprezíveis para k maior do que 5 ou 6.

A distribuição de Poisson é largamente empregada quando se deseja contar o número de eventos de certo tipo que ocorrem num intervalo de tempo, ou superfície ou volume. São exemplos:

(a) número de chamadas recebidas por um telefone durante cinco minutos;
(b) número de falhas de um computador num dia de operação; e
(c) número de relatórios de acidentes enviados a uma companhia de seguros numa semana.

De modo geral, dizemos que a v.a. N tem uma distribuição de Poisson com parâmetro $\lambda > 0$ se

$$P(N = k) = \frac{e^{-\lambda}\lambda^k}{k!},\quad k = 0, 1, 2, \ldots \tag{6.24}$$

É fácil verificar que $E(N) = \mathrm{Var}(N) = \lambda$ (veja o Problema 46); logo, λ representa o número médio de eventos ocorrendo no intervalo considerado.

Uma suposição que se faz usualmente em relação à distribuição de Poisson é que a probabilidade de se obter mais de um evento num intervalo muito pequeno é desprezível.

Exemplo 6.17. Uma situação prática de interesse na qual a distribuição de Poisson é empregada diz respeito à desintegração de substâncias radioativas. Considere o urânio 238 (U^{238}), por exemplo. Cada núcleo de U^{238} tem uma probabilidade muito pequena, $4{,}9 \times 10^{-18}$ de se desintegrar, emitindo uma partícula α, em um segundo. Considere, agora, um número grande n de núcleos e a v.a. N = número de núcleos que se desintegram. Admitindo-se que a desintegração de um núcleo não afeta a probabilidade de desintegração de qualquer outro núcleo (independência), a v.a. N tem uma distribuição binomial, com parâmetros n e p, este dado pelo valor acima. Logo, estamos numa situação em que podemos usar (6.23), ou seja, aproximar probabilidades binomiais por probabilidades de Poisson.

Em 0,30 mg de U^{238} temos aproximadamente $n = 7{,}6 \times 10^{17}$ átomos (Helene e Vanin, 1981), logo $\lambda = np \approx 3{,}7$ e

$$P(N = k) \approx \frac{e^{-3,7}(3{,}7)^k}{k!},\ k = 0, 1, \ldots$$

Por exemplo, $P(N = 0) = 0{,}025$ e $P(N = 2) = 0{,}169$. Pode-se ver que $P(N \geqslant 19)$ é muito pequena, menor do que 10^{-6}.

Seria interessante avaliar se a distribuição de Poisson realmente é um modelo razoável para essa situação. Um experimento devido a Rutherford e Geiger (veja Feller, 1964, pág. 149, para a referência completa sobre esse experimento) de fato comprova essa adequação. Eles observaram os números de partículas α emitidas por uma substância radioativa em $n = 2.608$ intervalos de 7,5 segundos. A Tabela 6.13 apresenta os números n_k de intervalos de 7,5 segundos contendo k partículas. Uma estimativa de λ = número médio de partículas emitidas durante um intervalo de 7,5 segundos é dada por

$$\lambda = \frac{\sum k n_k}{n} = \frac{10.094}{2.608} = 3{,}87.$$

As probabilidades de Poisson são dadas por

$$p_k = \frac{3{,}87^k e^{-3{,}87}}{k!}, \quad k = 0, 1, 2, \ldots$$

Segue-se que np_k é o número esperado de intervalos contendo k partículas, e esses valores também estão apresentados na Tabela 6.13. Vemos que há uma boa coincidência entre os valores das duas colunas. Um teste formal pode ser feito para verificar a adequação da distribuição de Poisson. Veja o Capítulo 14, Exemplo 14.5.

Tabela 6.13: Freqüências observadas e esperadas para o Exemplo 6.17.

k	n_k	np_k
0	57	54,399
1	203	210,523
2	383	407,361
3	525	525,496
4	532	508,418
5	408	393,515
6	273	253,817
7	139	140,325
8	45	67,882
9	27	29,189
$\geqslant 10$	16	17,075
	2.608	2.608,000

Se considerarmos ocorrências de eventos em intervalos de tempo de comprimento t, no lugar de intervalo unitário de tempo, basta ajustar o parâmetro λ na fórmula (6.24). Vejamos um exemplo.

Exemplo 6.18. Um telefone recebe, em média, cinco chamadas por minuto. Supondo que a distribuição de Poisson seja adequada nessa situação, obter a probabilidade de que o telefone não receba chamadas durante um intervalo de um minuto.

Segue-se que $\lambda = 5$ e

$$P(N = 0) = \frac{5^0 e^{-5}}{0!} = e^{-5} = 0{,}0067.$$

Por outro lado, se quisermos a probabilidade de obter no máximo duas chamadas em quatro minutos, teremos $\lambda = 20$ chamadas em quatro minutos, logo

$$P(N \leqslant 2) = P(N = 0) + P(N = 1) + P(N = 2) = e^{-20}\,(1 + 20 + 200) = 221e^{-20},$$

que é um número muito próximo de zero.

Esse exemplo nos mostra que a probabilidade de k ocorrências num intervalo fixo de comprimento t pode ser escrita como

$$P(N = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \quad k = 0, 1, 2, \ldots, \tag{6.25}$$

onde λ representa o número médio de ocorrências naquele intervalo. Denotaremos uma v.a. N com distribuição de Poisson de parâmetro λ por

$$N \sim \text{Pois}(\lambda).$$

Apresentamos, na Tabela 6.14, um resumo das distribuições discretas estudadas neste capítulo. Para cada uma temos a fórmula que dá a probabilidade de assumir cada valor, os possíveis valores, os parâmetros que caracterizam cada distribuição, a média e a variância. Incluímos, também, a distribuição geométrica, tratada no Problema 55.

Tabela 6.14: Modelos para variáveis discretas.

Modelo	$P(X = x)$	Parâmetros	$E(X)$, $\text{Var}(X)$
Bernoulli	$p^x(1-p)^{1-x}$, $x = 0, 1$	p	p, $p(1-p)$
Binomial	$\binom{n}{x}p^x(1-p)^{n-x}$, $x = 0, \ldots, n$	n, p	np, $np(1-p)$
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}$, $x = 0, 1, \ldots$	λ	λ, λ
Geométrica	$p(1-p)^{x-1}$, $x = 1, 2, \ldots$	p	$\frac{1}{p}$, $\frac{(1-p)}{p^2}$
Hipergeométrica	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$, $a \leqslant x \leqslant b^{(1)}$.	N, r, n	$\frac{nr}{N}$, $n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\frac{(N-n)}{(N-1)}$

(1) $a = \max(0, n - N + r)$, $b = \min(r, n)$.

Problemas

20. Para os exercícios *(a)* a *(e)* abaixo, considere o enunciado:
    Das variáveis abaixo descritas, assinale quais são binomiais, e para essas dê os respectivos campos de definição e função de probabilidade. Quando julgar que a variável não é binomial, aponte as razões de sua conclusão.

(a) De uma urna com dez bolas brancas e 20 pretas, vamos extrair, com reposição, cinco bolas. X é o número de bolas brancas nas cinco extrações.
(b) Refaça o problema anterior, mas dessa vez as n extrações são sem reposição.
(c) Temos cinco urnas com bolas pretas e brancas e vamos extrair uma bola de cada urna. Suponha que X seja o número de bolas brancas obtidas no final.
(d) Vamos realizar uma pesquisa em dez cidades brasileiras, escolhendo ao acaso um habitante de cada uma delas e classificando-o em pró ou contra um certo projeto federal. Suponha que X seja o número de indivíduos contra o projeto no final da pesquisa.
(e) Em uma indústria existem 100 máquinas que fabricam determinada peça. Cada peça é classificada como boa ou defeituosa. Escolhemos ao acaso um instante de tempo e verificamos uma peça de cada uma das máquinas. Suponha que X seja o número de peças defeituosas.

21. Se $X \sim b(n, p)$, sabendo-se que $E(X) = 12$ e $\sigma^2 = 3$, determinar:
(a) n
(b) p
(c) $P(X < 12)$
(d) $P(X \geqslant 14)$
(e) $E(Z)$ e $\text{Var}(Z)$, onde $Z = (X - 12)/\sqrt{3}$
(f) $P(Y \geqslant 14/16)$, onde $Y = X/n$
(g) $P(Y \geqslant 12/16)$, onde $Y = X/n$

22. Numa central telefônica, o número de chamadas chega segundo uma distribuição de Poisson, com a média de oito chamadas por minuto. Determinar qual a probabilidade de que num minuto se tenha:
(a) dez ou mais chamadas;
(b) menos que nove chamadas;
(c) entre sete (inclusive) e nove (exclusive) chamadas.

23. Num certo tipo de fabricação de fita magnética, ocorrem cortes a uma taxa de um por 2.000 pés. Qual a probabilidade de que um rolo com 2.000 pés de fita magnética tenha:
(a) nenhum corte?
(b) no máximo dois cortes?
(c) pelo menos dois cortes?

24. Suponha que a probabilidade de que um item produzido por uma máquina seja defeituoso é de 0,2. Se dez itens produzidos por essa máquina são selecionados ao acaso, qual é a probabilidade de que não mais do que um defeituoso seja encontrado? Use a binomial e a distribuição de Poisson e compare os resultados.

25. Examinaram-se 2.000 ninhadas de cinco porcos cada uma, segundo o número de machos. Os dados estão representados na tabela abaixo.

Nº de Machos	Nº de Ninhadas
0	20
1	360
2	700
3	680
4	200
5	40
Total	2.000

(a) Calcule a proporção média de machos.

(b) Calcule, para cada valor de X, o número de ninhadas que você deve esperar se $X \sim b(5, p)$, onde p é a proporção média de machos calculada em (a).

26. Se X tem distribuição binomial com parâmetros $n = 5$ e $p = 1/2$, faça os gráficos da distribuição de X e da f.d.a. $F(x)$.

27. Considere, agora, $n = 5$ e $p = 1/4$. Obtenha o gráfico da distribuição de X. Qual a diferença entre esse gráfico e o correspondente do Problema 26? O que ocasionou a diferença?

28. Refaça o Problema 26, com $n = 6$ e $p = 1/2$.

6.7 O Processo de Poisson

No Exemplo 6.17 acima vimos uma aplicação importante da distribuição de Poisson ao problema da desintegração radioativa. Lá tratamos da emissão de partículas alfa em intervalos de 7,5 segundos. Ou seja, estamos contando o número de ocorrências de um evento ao longo do tempo. Na realidade, consideramos o que se chama um *processo estocástico*. Designando-se por N_t o número de partículas emitidas no intervalo $[0, t)$, obteremos o que se chama de *processo de Poisson*, para todo $t \geqslant 0$. Nesta seção iremos partir de algumas suposições que consideramos plausíveis sobre tal processo e mostrar que a distribuição da variável aleatória N_t, para cada $t = 0$, é dada pela fórmula (6.25).

As suposições que iremos admitir como válidas são as seguintes.

(S1) $N_0 = 0$, ou seja, o processo começa no instante zero com probabilidade um: $P(N_0 = 0) = 1$.

(S2) Os números de eventos em intervalos de tempo disjuntos são v.a. independentes. Considere $0 < t < t + s$, N_t como antes e $N_{t+s} - N_t$ o número de eventos no intervalo $[t, t + s)$. Então, estamos supondo que as v.a. N_t e $N_{t+s} - N_t$ são independentes. Dizemos que o processo tem *incrementos independentes*.

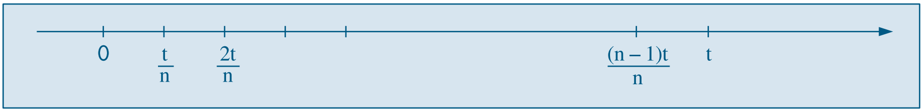
(S3) Considere os intervalos $[0, t)$ e $[s, s + t)$, de mesmo comprimento t e as v.a. N_t como antes e M_t = número de eventos no intervalo $[s, s + t)$. Então, para todo $s > 0$, as v.a. N_t e M_t têm a mesma distribuição de probabilidades. Ou seja, a distribuição do número de eventos ocorridos num intervalo depende somente do comprimento do intervalo, e não de sua localização. Dizemos que o processo tem *incrementos estacionários*.

(S4) Para h suficientemente pequeno, $P(N_h = 1) \approx \lambda h$, com $\lambda > 0$, constante. Ou seja, num intervalo pequeno, a probabilidade de ocorrência de um evento é proporcional ao comprimento do intervalo.

(S5) Para h como em (S4), $P(N_h \geqslant 2) \approx 0$. Isso nos diz que a probabilidade de se ter dois ou mais eventos num intervalo suficientemente pequeno é desprezível.

Considere o intervalo $[0, t)$ e o divida em subintervalos de comprimento t/n, como na Figura 6.13.

Figura 6.13: Divisão de intervalo $[0, t)$ em subintervalos de comprimentos t/n.



Chamemos de Y a v.a. que dá os números de subintervalos com um evento. Então, Y é uma v.a. com distribuição binomial, de parâmetros n (número total de subintervalos) e $p = P$ (um evento) $= \lambda(t/n)$. Para n grande, usando a aproximação da seção anterior, temos que essa variável pode ser aproximada por uma v.a. com distribuição de Poisson com parâmetro $np = n\lambda(t/n) = \lambda t$. Note que aqui usamos as suposições S2 (cada subintervalo contém um evento, independentemente dos demais intervalos) e S3 (com a mesma probabilidade).

Pela suposição S5, a probabilidade de que cada subintervalo contenha dois ou mais eventos tende a zero, quando n cresce. Logo, N_t é uma v.a. com distribuição de Poisson, com parâmetro λt.

Uma prova um pouco mais rigorosa, usando derivadas, pode ser dada. Veja Meyer (1965).

6.8 Quantis

No Capítulo 3 estudamos os quantis associados a um conjunto de dados. Esses poderiam ser chamados de quantis empíricos, pois podemos agora considerar quantis associados à distribuição de uma v.a. discreta, os quais poderíamos denominar quantis teóricos.

Definição. O valor $Q(p)$ satisfazendo

$$P(X \leqslant Q(p)) \geqslant p \text{ e } P(X \geqslant Q(p)) \geqslant 1 - p, \tag{6.26}$$

para $0 < p < 1$, é chamado o *p-quantil* de X.

A interpretação do p-quantil é similar à que foi dada no caso de um conjunto de dados: $Q(p)$ é o valor tal que a soma das probabilidades dos valores menores do que ele, é p. Então, por que não defini-lo por $F(Q(p)) = P(X \leqslant Q(p)) = p$, onde $F(x)$ é a f.d.a. de X? A resposta será dada acompanhando os exemplos a seguir.

Para determinados valores de p teremos, como antes, denominações especiais. Por exemplo:

$Q_1 = Q(0{,}25)$: primeiro quartil
$Q_2 = Q(0{,}5)$: mediana ou segundo quartil
$Q_3 = Q(0{,}75)$: terceiro quartil.

Vejamos o caso da mediana, $Q(0{,}5) = Md$. Por (6.26) devemos ter

$$P(X \leqslant Md) \geqslant 0{,}5 \text{ e } P(X \geqslant Md) \geqslant 0{,}5. \tag{6.27}$$

Suponha a v.a. X com a distribuição:

x	0	1
$p(x)$	1/3	2/3

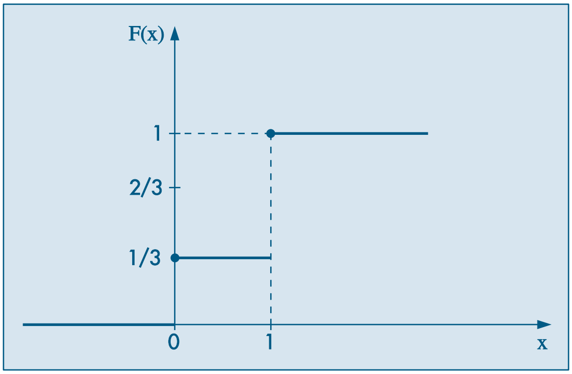
Então $Md = 1$, pois $P(X \leqslant 1) = 1/3 + 2/3 = 1 > 1/2$ e $P(X \geqslant 1) = P(X = 1) = 2/3 > 1/2$.

Na Figura 6.14 temos a f.d.a. de X. Sabemos que

$$F(x) = \begin{cases} 0, & x < 0 \\ 1/3, & 0 \leqslant x < 1 \\ 1, & x \geqslant 1, \end{cases}$$

de modo que não existe algum valor x tal que $F(x) = 0{,}5$, o que ilustra por que não podemos definir a mediana por meio de $F(Md) = 0{,}5$.

Figura 6.14: f.d.a. da v.a. X



Por outro lado, considere a v.a. Y com a distribuição da tabela abaixo:

Y	–1	0	1
$p(y)$	1/4	1/4	1/2

Então, qualquer valor Md entre 0 e 1 é uma mediana, pois

$$P(Y \leqslant Md) = P(Y = -1) + P(Y = 0) = 1/2 \geqslant 1/2 \text{ e}$$
$$P(Y \geqslant Md) = P(Y = 1) = 1/2 \geqslant 1/2.$$

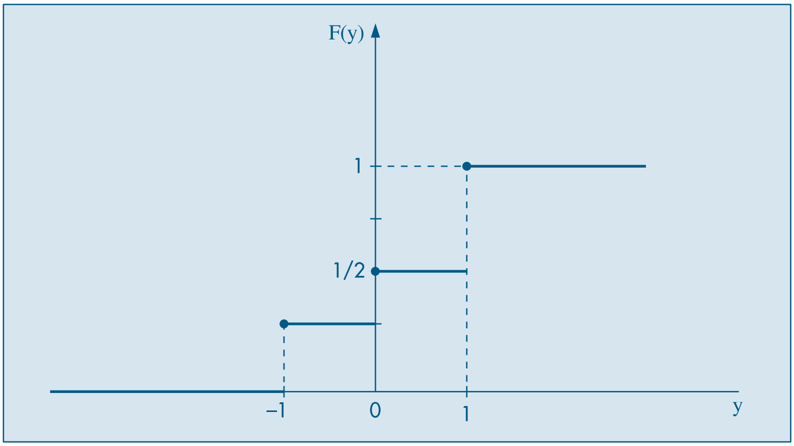
A f.d.a. de Y está na Figura 6.15. Observe que 0 e 1 também são medianas. Observe, também, que $Q(0{,}75) = 1$, pois

$$P(X \leqslant 1) = 1 \geqslant p = 0{,}75,$$

$$P(X \geqslant 1) = 0{,}5 \geqslant 1 - p = 0{,}25.$$

Novamente, não há nenhum valor de y tal que $F(y) = 0{,}75$. Mostre que $Q(0{,}90)$ também é igual a 1.

Figura 6.15: f.d.a. da v.a. Y



6.9 Exemplos Computacionais

Usando programas e planilhas computacionais é possível gerar probabilidades e probabilidades acumuladas para os modelos mais importantes discutidos neste capítulo. Por exemplo, o Minitab usa os comandos PDF para gerar probabilidades e CDF para gerar probabilidades acumuladas (f.d.a.).

Exemplo 6.19. Temos, no Quadro 6.1, as probabilidades $P(X = x)$ e $P(X \leqslant x)$ para uma v.a. $X \sim b(14; 0{,}3)$, ou seja, $n = 14$ e $p = P(\text{sucesso}) = 0{,}3$.

Quadro 6.1 Probabilidades binomiais geradas pelo Minitab.

MTB > PDF;
SUBC> Binomial 14 0.3.
Probability Density Function
Binomial with n = 14 and p = 0.300000

x	P(X = x)	x	P(X = x)
0	0.0068	7	0.0618
1	0.0407	8	0.0232
2	0.1134	9	0.0066
3	0.1943	10	0.0014
4	0.2290	11	0.0002
5	0.1963	12	0.0000
6	0.1262		

MTB > CDF;
SUBC> Binomial 14 0.3.
Cumulative Distribution Function
Binomial with n = 14 and p = 0.300000

x	P(X <= x)	x	P(X <= x)
0	0.0068	6	0.9067
1	0.0475	7	0.9685
2	0.1608	8	0.9917
3	0.3552	9	0.9983
4	0.5842	10	0.9998
5	0.7805	11	1.0000

Ainda, usando o Minitab, temos no Quadro 6.2 as probabilidades e probabilidades acumuladas para uma v.a. com distribuição de Poisson com parâmetro $\lambda = 5{,}2$.

Capítulo 7

Variáveis Aleatórias Contínuas

7.1 Introdução

Neste capítulo iremos estudar modelos probabilísticos para variáveis aleatórias contínuas, ou seja, variáveis para as quais os possíveis valores pertencem a um intervalo de números reais. A definição dada no capítulo anterior, para v.a. discreta, deve ser modificada como segue.

Definição. Uma função X, definida sobre o espaço amostral Ω e assumindo valores num intervalo de números reais, é dita uma *variável aleatória contínua*.

No Capítulo 2 vimos alguns exemplos de variáveis contínuas, como o salário de indivíduos, alturas etc. A característica principal de uma v.a. contínua é que, sendo resultado de uma mensuração, o seu valor pode ser pensado como pertencendo a um intervalo ao redor do valor efetivamente observado. Por exemplo, quando dizemos que a altura de uma pessoa é 175 cm, estamos medindo sua altura usando cm como unidade de medida e portanto o valor observado é, na realidade, um valor entre 174,5 cm e 175,5 cm.

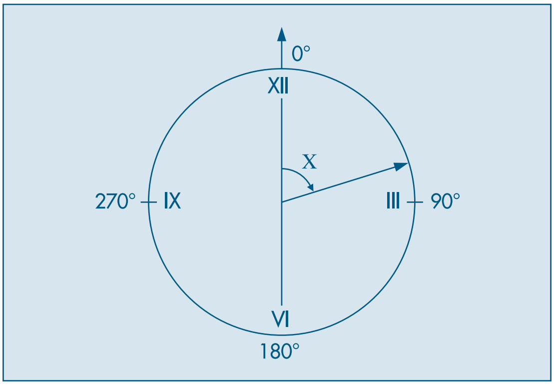
Vejamos um exemplo para motivar a discussão que se segue.

Exemplo 7.1. O ponteiro dos segundos de um relógio mecânico pode parar a qualquer instante, devido a algum defeito técnico, ou término da bateria, e vamos indicar por X o ângulo que esse ponteiro forma com o eixo imaginário passando pelo centro do mostrador e pelo número XII, conforme mostra a Figura 7.1.

Tabela 7.1: Distribuição uniforme discreta.

x	0°	6°	12°	18°	...	348°	354°
$p(x)$	1/60	1/60	1/60	1/60	...	1/60	1/60

Figura 7.1: Ilustração de uma v.a. X discreta.

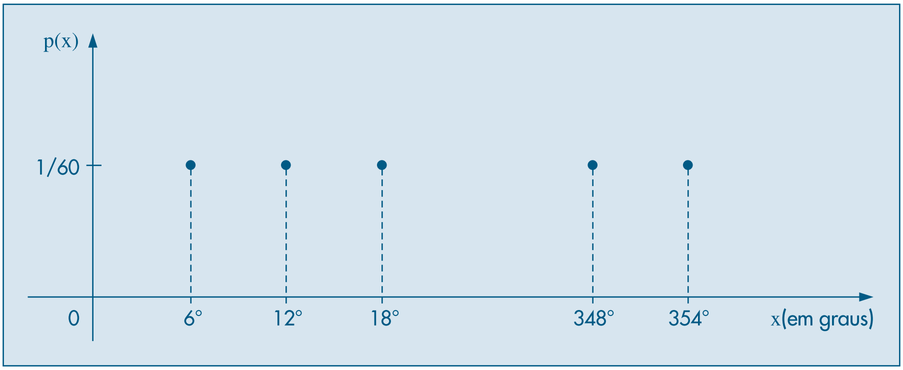


Medindo esse ângulo X em graus e lembrando que:

(i) o ponteiro deve dar 60 "saltos" (ele dá um salto em cada segundo) para completar uma volta;

(ii) acreditamos que o ponteiro tenha probabilidade igual de parar em qualquer ponto, então, a v.a. X tem distribuição uniforme discreta, com função de probabilidade dada pela Tabela 7.1 e representada graficamente na Figura 7.2.

Figura 7.2: Distribuição uniforme discreta.



Considerando esse mesmo problema com um relógio elétrico, para o qual o ponteiro dos segundos move-se *continuamente*, necessitamos de um outro modelo para representar a v.a. X. Primeiro, observamos que o conjunto dos possíveis valores de X não é mais um conjunto discreto de valores, pois X pode assumir qualquer valor do intervalo $[0,360) = = \{x \in \mathbb{R}: 0 \leqslant x < 360\}$. Em segundo lugar, como no caso do relógio mecânico, continuamos a acreditar que não exista uma região de preferência para o ponteiro parar. Como existem infinitos pontos nos quais o ponteiro pode parar, cada um com igual probabilidade, se fôssemos usar o mesmo método usado para a v.a. discreta uniforme, cada ponto teria probabilidade de ocorrer igual a zero. Assim não tem muito sentido falar na probabilidade de que o ângulo X seja igual a certo valor, pois essa probabilidade sempre será igual a zero. Entretanto, podemos determinar a probabilidade de que X esteja compreendido entre dois valores quaisquer. Por exemplo, usando a Figura 7.1 como referência, a probabilidade de o ponteiro parar no intervalo compreendido entre os números XII e III é 1/4, pois esse intervalo corresponde a 1/4 do intervalo total.

Podemos, pois, escrever

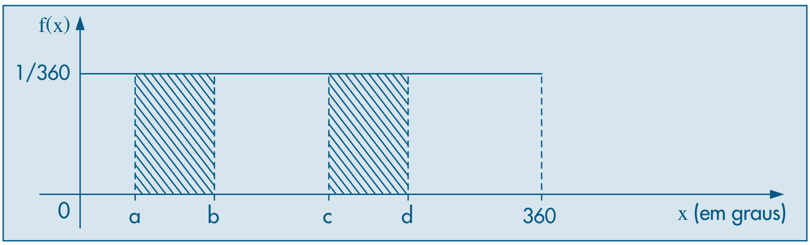
$$P(0^o \leqslant X \leqslant 90^o) = \frac{1}{4}.$$

Do mesmo modo, a probabilidade $P(120^o \leqslant X \leqslant 150^o) = 1/12$. Por menor que seja o intervalo, sempre poderemos calcular a probabilidade de o ponteiro parar num ponto qualquer desse intervalo. E é fácil verificar que, nesse caso, dados dois números a e b, tais que $0^o \leqslant a < b < 360^o$, a probabilidade de $X \in [a, b)$ é

$$P(a \leqslant X < b) = \frac{b-a}{360^o}$$

Através da divisão do intervalo $[0^o, 360^o)$ em pequenos subintervalos, podemos construir um histograma para as probabilidades da v.a. X (como fizemos para v.a contínuas no Capítulo 2). Ou ainda, como naquele capítulo, fazendo esses intervalos tenderem a zero, podemos construir o histograma alisado da v.a. X, apresentado na Figura 7.3.

Figura 7.3: Histograma alisado: distribuição uniforme contínua.



O histograma alisado da Figura 7.3 corresponde à seguinte função:

$$f(x) = \begin{cases} 0, & \text{se } x < 0^o \\ 1/360, & \text{se } 0^o \leqslant x < 360^o \\ 0, & \text{se } x \geqslant 360^o. \end{cases}$$

Como vimos na construção de histogramas, a área correspondente ao intervalo $[a, b)$ (hachurada na Figura 7.3) deve indicar a probabilidade de a variável estar entre a e b. Matematicamente, isso é expresso por meio da integral da função entre a e b; então,

$$P(a \leqslant X < b) = \int_a^b f(x)dx = \int_a^b \frac{1}{360}\, dx = \frac{b-a}{360},$$

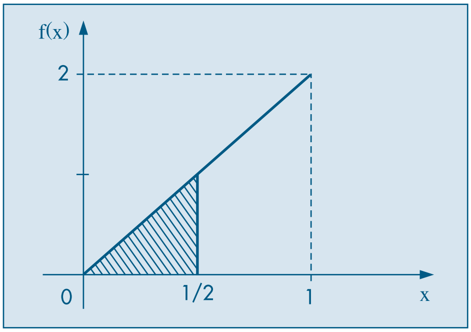
pois a integral definida de uma função entre dois pontos determina a área sob a curva representativa da função, compreendida entre esses dois pontos.

A função $f(x)$ é chamada *função densidade de probabilidade* (f.d.p.) da v.a. X.

Podemos construir modelos teóricos para variáveis aleatórias contínuas, escolhendo adequadamente as funções densidade de probabilidade. Teoricamente, qualquer função f, que seja não negativa e cuja área total sob a curva seja igual à unidade, caracterizará uma v.a. contínua.

Exemplo 7.2. Se $f(x) = 2x$, para $0 \leqslant x \leqslant 1$, e zero fora desse intervalo, vemos que $f(x) \geqslant 0$, para qualquer x, e a área sob o gráfico de $f(x)$ é unitária (verifique na Figura 7.4). Logo, a função f pode representar a função densidade de uma v.a. contínua X.

Figura 7.4: f.d.p. da v.a. X do Exemplo 7.2.



Para esse caso, $P(0 \leqslant X \leqslant 1/2)$ é igual à área do triângulo de base 1/2 e altura 1, hachurado na Figura 7.4; logo, a probabilidade em questão é

$$P(0 \leqslant X \leqslant 1/2) = \frac{1}{2}\left(\frac{1}{2} \times 1\right) = \frac{1}{4}.$$

Observamos, então, que a probabilidade de essa v.a. assumir um valor pertencente ao intervalo [0, 1/2) é menor que a probabilidade de a variável assumir um valor pertencente ao intervalo [1/2, 1).

A comparação das funções densidade dos dois últimos exemplos ajuda a entender seu significado. No primeiro exemplo, consideremos dois intervalos, $I_1 = [a, b)$ e $I_2 = [c, d)$, contidos no intervalo [0,360), com a mesma amplitude ($b - a = d - c$); então,

$$P(X \in I_1) = P(X \in I_2).$$

O mesmo não acontece no segundo exemplo: dados dois intervalos de mesma amplitude, aquele mais próximo de 1 irá apresentar maior probabilidade. Ou seja, a probabilidade de que a v.a. X assuma um valor num intervalo de amplitude fixa depende da posição do intervalo; existem regiões com maior *chance* de ocorrer, e o que determina esse fato é a função densidade de probabilidade. Portanto, a f.d.p. é um indicador da concentração de "massa" (probabilidade) nos possíveis valores de X. Convém ressaltar ainda que $f(x)$ não representa a probabilidade de ocorrência de algum evento. A área sob a curva entre dois pontos é que irá fornecer a probabilidade.

Problemas

1. Dada a função

$$f(x) = \begin{cases} 2e^{-2x}, & x \geqslant 0 \\ 0, & x < 0, \end{cases}$$

(a) Mostre que esta é uma f.d.p.
(b) Calcule a probabilidade de $X > 10$.

2. Uma v.a. X tem distribuição triangular no intervalo $[0, 1]$ se sua f.d.p. for dada por

$$f(x) = \begin{cases} 0, & x < 0 \\ Cx, & 0 \leqslant x \leqslant 1/2 \\ C(1-x), & 1/2 \leqslant x \leqslant 1 \\ 0, & x > 1. \end{cases}$$

(a) Qual valor deve ter a constante C?
(b) Faça o gráfico de $f(x)$.
(c) Determine $P(X \leqslant 1/2)$, $P(X > 1/2)$ e $P(1/4 \leqslant X \leqslant 3/4)$.

3. Suponha que estamos atirando dardos num alvo circular de raio 10 cm, e seja X a distância do ponto atingido pelo dardo ao centro do alvo. A f.d.p. de X é

$$f(x) = \begin{cases} kx, & \text{se } 0 \leqslant x \leqslant 10 \\ 0, & \text{para os demais valores.} \end{cases}$$

(a) Qual a probabilidade de acertar o centro do alvo, se esse for um círculo de 1 cm de raio?
(b) Mostre que a probabilidade de acertar qualquer círculo concêntrico é proporcional à sua área.

4. Encontre o valor da constante c se

$$f(x) = \begin{cases} c/x^2, & x \geqslant 10 \\ 0, & x < 10 \end{cases}$$

for uma densidade. Encontre $P(X > 15)$.

7.2 Valor Médio de uma Variável Aleatória Contínua

Do que foi visto até aqui, deduz-se que qualquer função $f(\cdot)$, não-negativa, tal que

$$\int_{-\infty}^{\infty} f(x)dx = 1,$$

define uma v.a. contínua X, ou seja, cria um modelo teórico para as freqüências relativas de uma v.a. contínua. A área compreendida entre dois valores, a e b, da abscissa x, sob a curva representativa de $f(x)$, dá a probabilidade (proporção teórica) da variável pertencer ao intervalo limitado pelos dois valores. Usando o conceito de integral, podemos escrever

$$P(a \leqslant X \leqslant b) = \int_a^b f(x)dx. \tag{7.1}$$

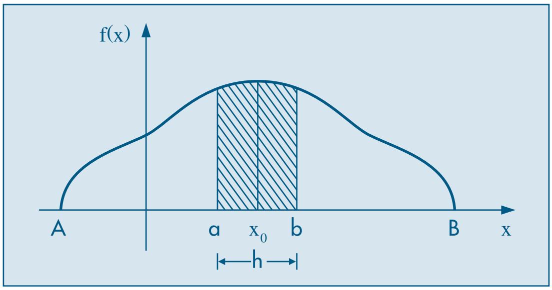
Vejamos agora como podemos definir a esperança (valor médio ou média) de uma v.a. contínua. Para isso, usaremos um artifício semelhante àquele usado na seção 3.1 para calcular a média das variáveis quantitativas, com os dados agrupados em classes. Lá substituímos todos os valores de um intervalo (classe) por um único valor aproximado (o ponto médio do intervalo), e agimos como se a variável fosse do tipo discreto. Aqui iremos repetir esse artifício.

Consideremos a v.a. X com função densidade $f(x)$ e dois pontos a e b, bem próximos, isto é, $h = b - a$ é pequeno, e consideremos x_0 o ponto médio do intervalo $[a, b]$. Observando a Figura 7.5 é fácil verificar que

$$P(a \leqslant X \leqslant b) \simeq h\, f(x_0), \tag{7.2}$$

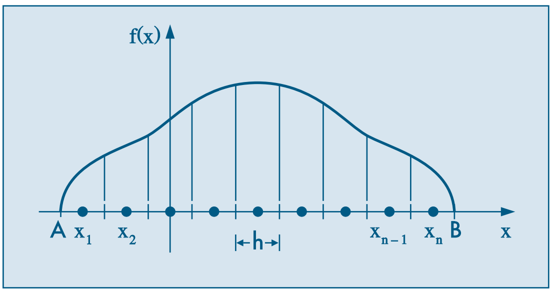
o que significa aproximar a área da parte hachurada pelo retângulo de base h e altura $f(x_0)$. É fácil ver que a aproximação melhora com h tendendo a zero.

Figura 7.5: Área hachurada representa $P(a \leqslant X \leqslant b)$.



Dividamos agora o intervalo $[A, B]$, onde $f(x) > 0$, em n partes de amplitudes iguais a $h = (B - A)/n$ (Figura 7.6) e consideremos os pontos médios desses intervalos, $x_1, x_2, ..., x_n$.

Figura 7.6: Partição do intervalo $[A, B]$.



Consideremos a v.a. Y_n, assumindo os valores $x_1, \ldots, x_n$ com as probabilidades

$$p_i = P(Y_n = x_i) \simeq f(x_i)h.$$

Dessa maneira, e de acordo com a definição de esperança, temos

$$E(Y_n) = \sum_{i=1}^{n} x_i p_i \approx \sum_{i=1}^{n} x_i\, f(x_i)h,$$

que será uma aproximação da esperança $E(X)$. Para determinar $E(X)$ com maior precisão, podemos aumentar o número de intervalos, diminuindo sua amplitude h. No limite, quando $h \to 0$, teremos o valor de $E(X)$. Definamos, pois,

$$E(X) = \lim_{n\to\infty} E(Y_n) = \lim_{n\to\infty} \sum_{i=1}^{n} x_i\, f(x_i)h. \tag{7.3}$$

Mas da definição de integral (veja Morettin *et al.*, 2005), temos que, se o limite (7.3) existe, ele define a integral de $x\,f(x)$ entre A e B, isto é,

$$E(X) = \int_A^B x f(x)dx. \tag{7.4}$$

Exemplo 7.3. Continuando com o Exemplo 7.2, observamos que, dividindo o intervalo $[0, 1]$ em n subintervalos, teremos $h = 1/n$, $x_i = (2i - 1)/2n$ e $f(x_i) = (2i - 1)/n$, $i = 1, 2, \ldots, n$. Portanto,

$$E(Y_n) = \sum_{i=1}^{n} \left(\frac{2i-1}{2\,n}\right)\left(\frac{2i-1}{n}\right)\left(\frac{1}{n}\right) = \frac{1}{2n^3}\sum_{i=1}^{n}(2i-1)^2$$

$$= \frac{1}{2n^3}\left\{\frac{n(2n+1)(2n-1)}{3}\right\} = \frac{1}{6}\left(2+\frac{1}{n}\right)\left(2-\frac{1}{n}\right),$$

na qual usamos o conhecido resultado que dá a soma dos quadrados dos primeiros n números ímpares. Logo,

$$E(X) = \lim_{n\to\infty} \frac{1}{6}\left(2+\frac{1}{n}\right)\left(2-\frac{1}{n}\right) = \frac{2}{3}.$$

O mesmo resultado é obtido diretamente da relação (7.4):

$$E(X) = \int_0^1 (x)(2x)dx = \left[\frac{2x^3}{3}\right]_0^1 = \frac{2}{3}.$$

Exemplo 7.4. No caso do relógio elétrico do Exemplo 7.1, obtemos

$$E(X) = \int_0^{360} x\,\frac{1}{360}\,dx = \left[\frac{1}{360}\,\frac{x^2}{2}\right]_0^{360} = 180,$$

que é o valor esperado devido à distribuição uniforme das freqüências teóricas.

Como a função $f(x)$ é sempre não-negativa, podemos escrever a esperança como

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx. \tag{7.5}$$

A extensão do conceito de variância para v.a. contínuas é feita de maneira semelhante e o equivalente à expressão (6.2) é

$$\text{Var}(X) = E[(X - E(X))^2] = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx. \tag{7.6}$$

Exemplo 7.5. Para os dois exemplos vistos anteriormente, teremos:

(i) Para o caso do relógio,

$$\text{Var}(X) = \int_0^{360} (x - 180)^2 \frac{1}{360} dx = \frac{1}{360} \left[\frac{x^3}{3} - \frac{360x^2}{2} + 180^2 x \right]_0^{360} = 10.800;$$

(ii) Para o Exemplo 7.2,

$$\text{Var}(X) = \int_0^1 \left(x - \frac{2}{3} \right)^2 2x dx = 2\left[\frac{x^4}{4} - \frac{4x^3}{9} + \frac{2x^2}{9} \right]_0^1 = \frac{1}{18}.$$

Como no caso de v.a. discretas, o desvio padrão de uma v.a. contínua X é definido como

$$DP(X) = \sqrt{\text{Var}(X)}, \tag{7.7}$$

que é dado na mesma unidade de medida do que X. Deixamos a cargo do leitor a verificação de que o seguinte resultado vale, como conseqüência de (7.6):

$$\text{Var}(X) = E(X^2) - [E(X)]^2. \tag{7.8}$$

Como frisamos no Capítulo 6, freqüentemente usaremos outros símbolos para indicar os parâmetros discutidos, a saber:

$$E(X) = \mu(X),$$
$$\text{Var}(X) = \sigma^2(X),$$
$$DP(X) = \sigma(X),$$

ou simplesmente μ, σ^2 e σ, respectivamente, se não houver possibilidade de confusão.

7.3 Função de Distribuição Acumulada

Dada uma v.a. X com função densidade de probabilidade $f(x)$, podemos definir a sua função de distribuição acumulada, $F(x)$, do mesmo modo como foi definida no Capítulo 6:

$$F(x) = P(X \leqslant x), -\infty < x < \infty. \tag{7.9}$$

De (7.1) segue-se que

$$F(x) = \int_{-\infty}^{x} f(t) dt, \tag{7.10}$$

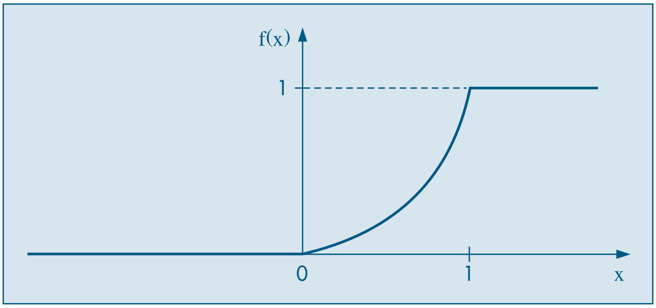
para todo real x.

Exemplo 7.6. Retomemos o Exemplo 7.2. Temos

$$F(x) = \begin{cases} 0, & \text{se } x < 0 \\ \int_0^x 2t dt = x^2, & \text{se } 0 \leqslant x < 1 \\ \int_0^1 2t dt + \int_1^x 0 dt = 1, & \text{se } x \geqslant 1. \end{cases}$$

O gráfico de $F(x)$ está na Figura 7.7.

Figura 7.7: f.d.a. da v.a. X do Exemplo 7.6.



De (7.9), vemos que $0 \leqslant F(x) \leqslant 1$, para todo x real; além disso, $F(x)$ é não-decrescente e possui as duas seguintes propriedades:

(i) $\lim_{x \to -\infty} F(x) = 0$,

(ii) $\lim_{x \to \infty} F(x) = 1$.

No Exemplo 7.6 temos, efetivamente, $F(x) = 0$, para $x < 0$ e $F(x) = 1$, para $x \geqslant 1$.

Para v.a. contínuas, o seguinte resultado é importante.

Proposição 7.1. Para todos os valores de x para os quais $F(x)$ é derivável temos

$$F'(x) = \frac{dF(x)}{dx} = f(x).$$

Vamos usar esse resultado no exemplo a seguir.

Exemplo 7.7. Suponha que

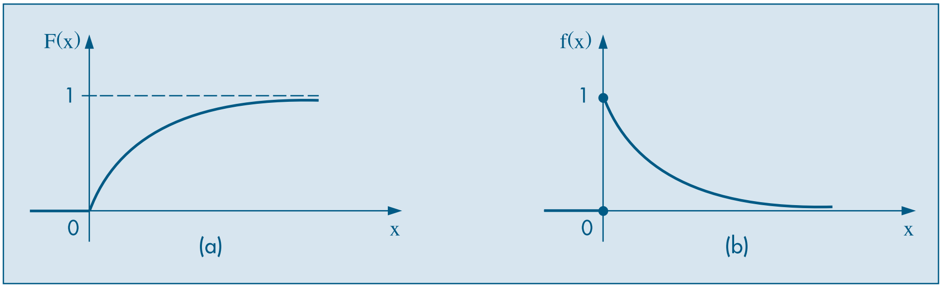
$$F(x) = \begin{cases} 0, & \text{se } x < 0 \\ 1 - e^{-x}, & \text{se } x \geqslant 0 \end{cases}$$

seja a f.d.a. de uma v.a. X. Então,

$$f(x) = \begin{cases} 0, & \text{se } x < 0 \\ e^{-x}, & \text{se } x \geqslant 0. \end{cases}$$

Na Figura 7.8 temos os gráficos dessas duas funções. Veremos que f(x) é um caso especial da densidade exponencial, a ser estudada na seção 7.4.3.

Figura 7.8: Distribuição exponencial ($\beta = 1$) (a) f.d.a. (b) f.d.p.



Se a e b forem dois números reais quaisquer,

$$P(a < X \leqslant b) = F(b) - F(a). \tag{7.11}$$

Esse resultado não será afetado se incluirmos ou não os extremos a e b na desigualdade entre parênteses.

Problemas

5. Calcule a esperança, a variância e a f.d.a. da v.a. X do Problema 2.

6. Determine a esperança e a variância da v.a. cuja f.d.p. é

$$f(x) = \begin{cases} \text{sen}\, x, & 0 \leqslant x \leqslant \pi/2 \\ 0, & \text{caso contrário.} \end{cases}$$

7. Calcule a média da v.a. X do Problema 4.

8. A v.a. contínua X tem f.d.p.

$$f(x) = \begin{cases} 3x^2, & -1 \leqslant x \leqslant 0 \\ 0, & \text{caso contrário.} \end{cases}$$

   (a) Se b for um número que satisfaz $-1 < b < 0$, calcule $P(X > b \mid X < b/2)$.
   (b) Calcule $E(X)$ e $\text{Var}(X)$.

9. Certa liga é formada pela mistura fundida de dois metais. A liga resultante contém certa porcentagem de chumbo, X, que pode ser considerada uma v.a. com f.d.p.

$$f(x) = \frac{3}{5}\, 10^{-5}\, x(100 - x),\ 0 \leqslant x \leqslant 100.$$

   Suponha que L, o lucro líquido obtido na venda dessa liga (por unidade de peso), seja dado por $L = C_1 + C_2X$. Calcule $E(L)$, o lucro esperado por unidade.

10. A demanda diária de arroz num supermercado, em centenas de quilos, é uma v.a. com f.d.p.

$$f(x) = \begin{cases} 2x/3, & \text{se } 0 \leqslant x < 1 \\ -x/3 + 1, & \text{se } 1 \leqslant x < 3 \\ 0, & \text{se } x < 0 \text{ ou } x > 3. \end{cases}$$

   (a) Qual a probabilidade de se vender mais do que 150 kg, num dia escolhido ao acaso?
   (b) Em 30 dias, quanto o gerente do supermercado espera vender?
   (c) Qual a quantidade de arroz que deve ser deixada à disposição dos clientes diariamente para que não falte arroz em 95% dos dias?

11. Suponha que X tenha f.d.p. $f(x)$ do Problema 1. Calcule $E(X)$ e $\text{Var}(X)$.

12. Seja X com densidade

$$f(x) = \begin{cases} c(1 - x^2), & \text{se } -1 \leqslant x \leqslant 1 \\ 0, & \text{caso contrário.} \end{cases}$$

   Calcule a média e a variância de X.

7.4 Alguns Modelos Probabilísticos para Variáveis Aleatórias Contínuas

De modo geral, podemos dizer que as v.a. cujos valores resultam de algum processo de mensuração são v.a. contínuas. Alguns exemplos são:

(a) o peso ou a altura das pessoas de uma cidade;

(b) a demanda diária de arroz num supermercado;

(c) o tempo de vida de uma lâmpada;

(d) o diâmetro de rolamentos de esferas; e

(e) erros de medidas em geral, resultantes de experimentos em laboratórios.

Dada uma v.a. contínua X, interessa saber qual a f.d.p. de X. Alguns modelos são freqüentemente usados para representar a f.d.p. de v.a. contínuas. Alguns dos mais utilizados serão descritos a seguir e, para uniformizar o estudo desses modelos, iremos em cada caso analisar:

(a) definição;

(b) gráfico da f.d.p.;

(c) momentos: $E(X)$, $\text{Var}(X)$;

(d) função de distribuição acumulada (f.d.a.).

Outros modelos serão apresentados na seção 7.7.

7.4.1 O Modelo Uniforme

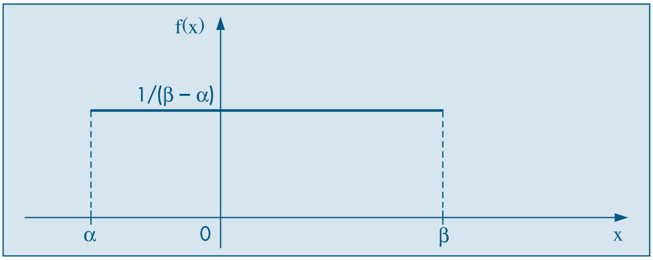
O modelo uniforme é uma generalização do modelo estudado no Exemplo 7.1 e é o modelo mais simples para v.a. contínuas.

(a) **Definição.** A v.a. X tem distribuição uniforme no intervalo $[\alpha, \beta]$ se sua f.d.p. é dada por

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{1}{\beta - \alpha}, & \text{se } \alpha \leqslant x \leqslant \beta, \\ 0, & \text{caso contrário.} \end{cases} \tag{7.12}$$

(b) **Gráfico.** A Figura 7.9 representa a função dada por (7.12).

Figura 7.9: Distribuição uniforme no intervalo $[\alpha, \beta]$.



(c) **Momentos.** Pode-se mostrar (veja o Problema 29) que

$$E(X) = \frac{\alpha + \beta}{2}, \tag{7.13}$$

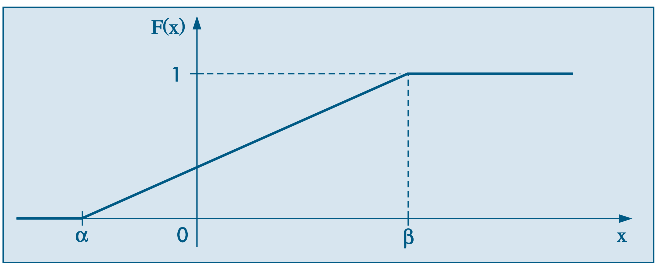
$$\text{Var}(X) = \frac{(\beta - \alpha)^2}{12}. \tag{7.14}$$

(d) **F.d.a.** A função de distribuição acumulada da uniforme é fácil de ser encontrada (veja o Problema 29):

$$F(x) = P(X \leqslant x) = \int_{-\infty}^{x} f(x)dx = \begin{cases} 0, & \text{se } x < \alpha \\ \dfrac{x - \alpha}{\beta - \alpha}, & \text{se } \alpha \leqslant x < \beta \\ 1, & \text{se } x \geqslant \beta, \end{cases} \tag{7.15}$$

cujo gráfico está na Figura 7.10.

Figura 7.10: f.d.a. de uma v.a. uniforme no intervalo $[\alpha, \beta]$.



Assim, para dois valores quaisquer c e d, $c < d$, teremos

$$P(c < X \leqslant d) = F(d) - F(c),$$

que é obtida facilmente de (7.15).

Usaremos a notação

$$X \sim \mathcal{U}(\alpha, \beta)$$

para indicar que a v.a. X tem distribuição uniforme no intervalo $[\alpha, \beta]$.

Exemplo 7.8. Um caso particular bastante interessante é aquele em que $\alpha = -1/2$ e $\beta = 1/2$. Indicando essa v.a. por U, teremos

$$f(u) = \begin{cases} 1, & \text{se } -1/2 \leqslant u \leqslant 1/2 \\ 0, & \text{caso contrário.} \end{cases}$$

Nessa situação temos que

$$E(U) = 0, \ \text{Var}(U) = 1/12$$

e a f.d.a. é dada por

$$F_U(u) = \begin{cases} 0, & \text{se } u < -1/2 \\ u + 1/2, & \text{se } -1/2 \leqslant u < 1/2 \\ 1, & \text{se } u > 1/2. \end{cases}$$

Por exemplo,

$$P(-1/4 \leqslant U \leqslant 1/4) = F_U(1/4) - F_U(-1/4) = 1/2.$$

Se quiséssemos facilitar o nosso trabalho, poderíamos tabelar os valores da f.d.a para essa variável U. Devido à simetria da área em relação a $x = 0$, poderíamos construir uma tabela indicando a função $G(u)$, tal que

$$G(u) = P(0 \leqslant U \leqslant u)$$

para alguns valores de u (veja o Problema 30).

Dada uma v.a. uniforme X qualquer, com parâmetros α e β, podemos definir a v.a. U como

$$U = \frac{X - \dfrac{\beta+\alpha}{2}}{\beta - \alpha}. \tag{7.16}$$

Segue-se que a transformação (7.16) leva uma uniforme no intervalo $[\alpha, \beta]$ numa uniforme no intervalo $[-1/2, 1/2]$ e para dois números quaisquer c e d, com $c < d$,

$$P(c < X \leqslant d) = F(d) - F(c) = P\left(\frac{c - \dfrac{\beta+\alpha}{2}}{\beta-\alpha} < U \leqslant \frac{d - \dfrac{\beta+\alpha}{2}}{\beta-\alpha}\right) = F_U\left(\frac{d - \dfrac{\beta+\alpha}{2}}{\beta-\alpha}\right) - F_U\left(\frac{c - \dfrac{\beta+\alpha}{2}}{\beta-\alpha}\right).$$

Artifícios semelhantes a esse são muito úteis na construção de tabelas e programas para cálculos de probabilidades referentes a famílias de modelos.

Um outro caso importante é para $\alpha = 0$ e $\beta = 1$. Um número aleatório é um valor gerado de uma v.a. com distribuição uniforme no intervalo [0, 1]. Veja Capítulo 9.

7.4.2 O Modelo Normal

Vamos introduzir, agora, um modelo fundamental em probabilidades e inferência estatística. Suas origens remontam a Gauss em seus trabalhos sobre erros de observações astronômicas, por volta de 1810, donde o nome de distribuição *gaussiana* para tal modelo.

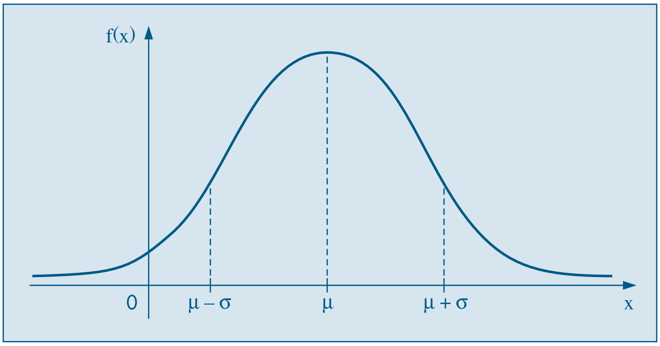
(a) **Definição.** Dizemos que a v.a. X tem *distribuição normal* com parâmetros μ e σ^2, $-\infty < \mu < +\infty$ e $0 < \sigma^2 < \infty$, se sua densidade é dada por

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty. \tag{7.17}$$

Claramente, $f(x; \mu, \sigma^2) \geqslant 0$, para todo x e pode-se provar que $\int_{-\infty}^{\infty} f(x; \mu, \sigma^2)\, dx = 1$. Veja o Problema 60.

(b) **Gráfico.** A Figura 7.11 ilustra uma particular *curva normal*, determinada por valores particulares de μ e σ^2.

Figura 7.11: f.d.p. de uma v.a. normal com média μ e desvio padrão σ.



(c) **Momentos.** Pode-se demonstrar que (veja o Problema 32):

$$E(X) = \mu, \tag{7.18}$$

$$\mathrm{Var}(X) = \sigma^2. \tag{7.19}$$

Além disso, $f(x; \mu; \sigma^2) \to 0$, quando $x \to \pm\infty$, $\mu - \sigma$ e $\mu + \sigma$ são pontos de inflexão de $f(x; \mu, \sigma^2)$, $x = \mu$ é ponto de máximo de $f(x; \mu, \sigma^2)$, e o valor máximo é $1/\sigma\sqrt{2\pi}$. A densidade $f(x; \mu, \sigma^2)$ é simétrica em relação à reta $x = \mu$, isto é,

$$f(\mu + x; \mu, \sigma^2) = f(\mu - x; \mu, \sigma^2), \tag{7.20}$$

para todo x real.

Para simplificar a notação, denotaremos a densidade da normal simplesmente por $f(x)$ e escreveremos, simbolicamente,

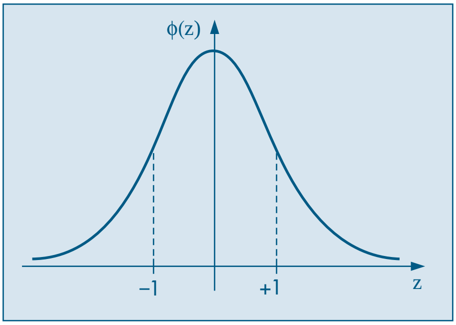
$$X \sim N(\mu, \sigma^2).$$

Quando $\mu = 0$ e $\sigma^2 = 1$, temos uma distribuição *padrão* ou *reduzida*, ou brevemente $N(0,1)$. Para essa a função densidade reduz-se a

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad -\infty < z < \infty. \tag{7.21}$$

O gráfico da normal padrão está na Figura 7.12.

Figura 7.12: f.d.p. de uma v.a. normal padrão: Z ~ N(0, 1).



Se $X \sim N(\mu;\ \sigma^2)$, então a v.a. definida por

$$Z = \frac{X-\mu}{\sigma}, \tag{7.22}$$

terá média zero e variância 1 (prove esses fatos). O que não é tão fácil mostrar é que Z também tem distribuição normal. Isso não será feito aqui.

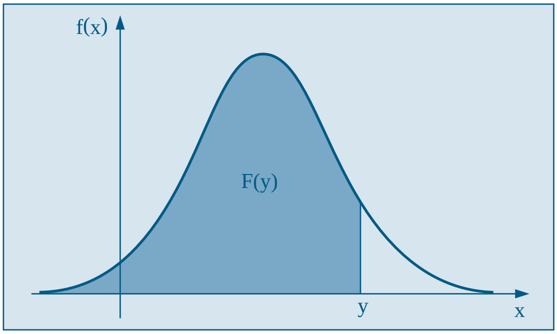
A transformação (7.22) é fundamental para calcularmos probabilidades relativas a uma distribuição normal qualquer.

(d) **F.d.a.** A f.d.a. $F(y)$ de uma v.a. normal X, com média μ e variância σ^2 é obtida integrando-se (7.17) de $-\infty$ até y, ou seja,

$$F(y) = \int_{-\infty}^{y} f(x;\ \mu,\ \sigma^2)dx, \quad y \in \mathbb{R}. \tag{7.23}$$

A integral (7.23) corresponde à área, sob $f(x)$, desde $-\infty$ até y, como ilustra a Figura 7.13.

Figura 7.13: Representação gráfica de F(y) como área.

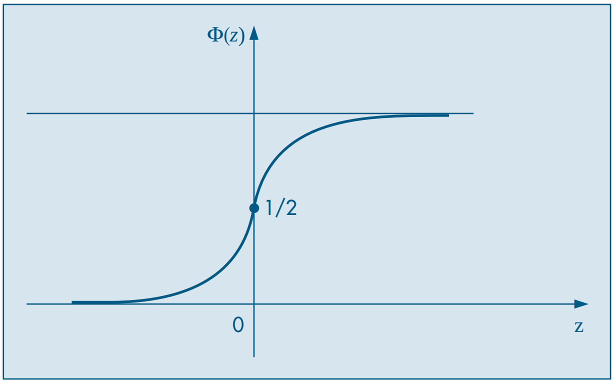


No caso específico da normal padrão, utilizamos a seguinte notação, que é universal:

$$\Phi(y) = \int_{-\infty}^{y} \phi(z)dz = 1/\sqrt{2\pi} \int_{-\infty}^{y} e^{-z^2/2}\, dz. \tag{7.24}$$

O gráfico de $\Phi(z)$ é ilustrado na Figura 7.14.

Figura 7.14: f.d.a. da normal padrão.

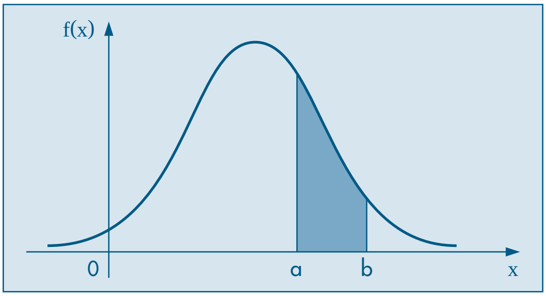


Suponha, então, que $X \sim N(\mu, \sigma^2)$ e que queiramos calcular

$$P(a < X < b) = \int_{a}^{b} f(x)dx, \tag{7.25}$$

onde $f(x)$ é dada por (7.17). Ver Figura 7.15.

Figura 7.15: Ilustração gráfica da $P(a \leqslant X \leqslant b)$ para uma v.a. normal.



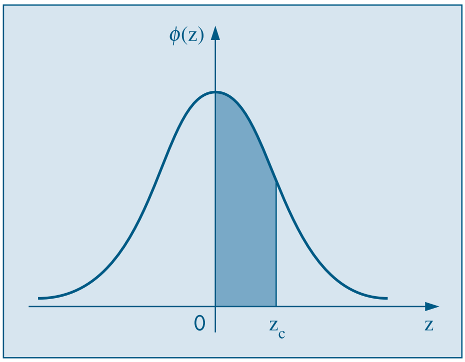
A integral (7.25) não pode ser calculada analiticamente, e portanto a probabilidade indicada só poderá ser obtida, aproximadamente, por meio de integração numérica. No entanto, para *cada* valor de μ e *cada* valor de σ, teríamos de obter $P(a < X < b)$ para diversos valores de a e b. Essa tarefa é facilitada através do uso de (7.22), de sorte que somente é necessário construir uma tabela para a distribuição normal padrão.

Vejamos, então, como obter probabilidades a partir da Tabela III. Essa tabela dá as probabilidades sob uma curva normal padrão, que nada mais são do que as correspondentes áreas sob a curva. A Figura 7.16 ilustra a probabilidade fornecida pela tabela, a saber,

$$P(0 \leqslant Z \leqslant z_c),$$

onde $Z \sim N(0,1)$.

Figura 7.16: $P(0 \leqslant Z \leqslant z_c)$ fornecido pela Tabela III.



Se tomarmos, por exemplo, $z_c = 1{,}73$, segue-se que

$$P(0 \leqslant Z \leqslant 1{,}73) = 0{,}4582.$$

Calculemos mais algumas probabilidades (Figura 7.17):

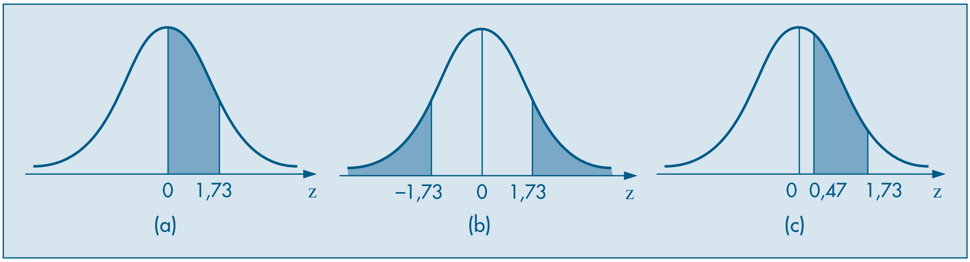
(a) $P(-1{,}73 \leqslant Z \leqslant 0) = P(0 \leqslant Z \leqslant 1{,}73) = 0{,}4582$, devido à simetria da curva.

(b) $P(Z \geqslant 1{,}73) = 0{,}5 - P(0 \leqslant Z \leqslant 1{,}73) = 0{,}5 - 0{,}4582 = 0{,}0418$, pois $P(Z \geqslant 0) = 0{,}5 = P(Z \leqslant 0)$.

(c) $P(Z < -1{,}73) = P(Z > 1{,}73) = 0{,}0418$.

(d) $P(0{,}47 \leqslant Z \leqslant 1{,}73) = P(0 \leqslant Z \leqslant 1{,}73) - P(0 \leqslant Z \leqslant 0{,}47) =$
$= 0{,}4582 - 0{,}1808 = 0{,}2774$.

Figura 7.17: Ilustração do cálculo de probabilidades para a $N(0{,}1)$.



Suponha, agora, que X seja uma v.a. $N(\mu, \sigma^2)$, com $\mu = 3$ e $\sigma^2 = 16$, e queiramos calcular $P(2 \leqslant X \leqslant 5)$. Utilizando (7.22), temos

$$P(2 \leqslant X \leqslant 5) = P\left(\frac{2-\mu}{\sigma} \leqslant \frac{X-\mu}{\sigma} \leqslant \frac{5-\mu}{\sigma}\right)$$

$$= P\left(\frac{2-3}{4} \leqslant Z \leqslant \frac{5-3}{4}\right) = P\left(-\frac{1}{4} \leqslant Z \leqslant \frac{1}{2}\right).$$

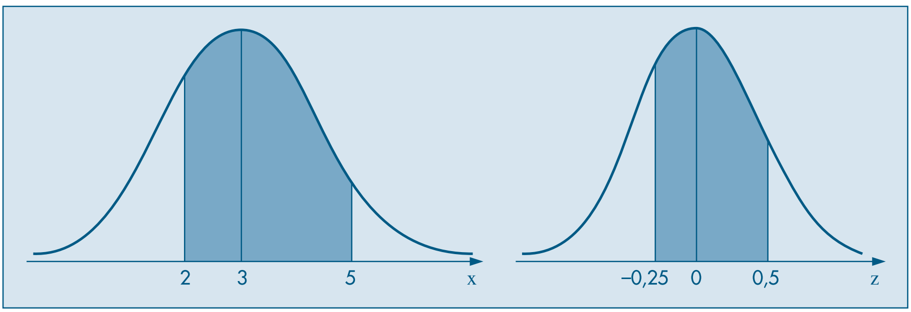
Portanto, a probabilidade de que X esteja entre 2 e 5 é igual à probabilidade de que Z esteja entre $-0{,}25$ e $0{,}5$ (Figura 7.18). Utilizando a Tabela III, vemos que

$$P(-0{,}25 \leqslant Z \leqslant 0{,}5) = 0{,}0987 + 0{,}1915 = 0{,}2902,$$

ou seja,

$$P(2 \leqslant X \leqslant 5) = 0{,}2902.$$

Figura 7.18: Ilustração do cálculo de $P(2 \leqslant X \leqslant 5)$ para a v.a. $N(3, 16)$.



Exemplo 7.9. Os depósitos efetuados no Banco da Ribeira durante o mês de janeiro são distribuídos normalmente, com média de \$10.000,00 e desvio padrão de \$1.500,00. Um depósito é selecionado ao acaso dentre todos os referentes ao mês em questão. Encontrar a probabilidade de que o depósito seja:

(a) \$10.000,00 ou menos;

(b) pelo menos \$10.000,00;

(c) um valor entre \$12.000,00 e \$15.000,00;

(d) maior do que \$20.000,00.

Temos que $\mu = 10.000$ e $\sigma = 1.500$. Seja a v.a. X = depósito.

(a) $P(X \leqslant 10.000) = P\left(Z \leqslant \dfrac{10.000 - 10.000}{1.500}\right) = P(Z \leqslant 0) = 0{,}5.$

(b) $P(X \geqslant 10.000) = P(Z \geqslant 0) = 0{,}5.$

(c) $P(12.000 < X < 15.000) = P\left(\dfrac{12.000 - 10.000}{1.500} < Z < \dfrac{15.000 - 10.000}{1.500}\right)$

$$= P(4/3 < Z < 10/3) = P(1{,}33 < Z < 3{,}33) = 0{,}09133.$$

(d) $P(X > 20.000) = P\left(Z > \dfrac{20.000 - 10.000}{1.500}\right) = P(Z > 6{,}67) \simeq 0.$

7.4.3 O Modelo Exponencial

Outra distribuição importante e que tem aplicações em confiabilidade de sistemas, assunto de que já tratamos brevemente no Capítulo 5, é a exponencial.

(a) **Definição.** A v.a. T tem *distribuição exponencial* com parâmetro $\beta > 0$ se sua f.d.p. tem a forma

$$f(t;\beta) = \begin{cases} \dfrac{1}{\beta}\, e^{-t/\beta}, & \text{se } t \geqslant 0 \\ 0, & \text{se } t < 0. \end{cases} \tag{7.26}$$

Escreveremos, brevemente,

$$T \sim \text{Exp}(\beta).$$

(b) **Gráfico.** O gráfico de $f(t;\beta) = f(t)$ está ilustrado na Figura 7.8 (b), com $\beta = 1$.

(c) **Momentos.** Usando integração por partes, pode-se demonstrar que (veja o Problema 41):

$$E(T) = \beta, \tag{7.27}$$

$$\text{Var}(T) = \beta^2. \tag{7.28}$$

Exemplo 7.10. O tempo de vida (em horas) de um transistor pode ser considerado uma v.a com distribuição exponencial com $\beta = 500$. Segue-se que a vida média do transistor é $E(T) = 500$ horas e a probabilidade de que ele dure mais do que a média é

$$P(T > 500) = \int_{500}^{\infty} f(t)dt = 1/500 \int_{500}^{\infty} e^{-t/500}\, dt$$

$$= 1/500\, [-500e^{-t/500}]_{500}^{\infty} = e^{-1} = 0{,}3678.$$

(d) **F.d.a.** Usando a definição (7.10), obtemos

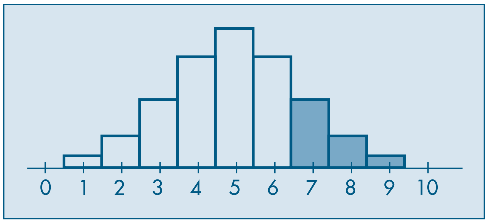
$$F(t) = \begin{cases} 0, & \text{se } t < 0 \\ 1 - e^{-t/\beta}, & \text{se } t \geqslant 0. \end{cases} \tag{7.29}$$

O gráfico de $F(t)$ está na Figura 7.8 (a), com $\beta = 1$.

7.5 Aproximação Normal à Binomial

Suponha que a v.a. Y tenha uma distribuição binomial com parâmetros $n = 10$ e $p = 1/2$ e queiramos calcular $P(Y \geqslant 7)$. Embora seja uma v.a. discreta, vimos no Capítulo 2 que é possível representá-la por meio de um histograma, como na Figura 7.19. Vemos que $P(Y = 7)$ é igual à área do retângulo de base unitária e altura igual a $P(Y = 7)$, similarmente para $P(Y = 8)$ etc. Logo, $P(Y \geqslant 7)$ é igual à soma das áreas dos retângulos hachurados na Figura 7.19.

Figura 7.19: $(P(Y \geqslant 7)$ para $Y \sim b(10, 1/2)$.



A idéia é aproximar tal área pela área sob uma curva normal, à *direita* de 6,5. Qual curva normal? Parece razoável considerar aquela normal de média

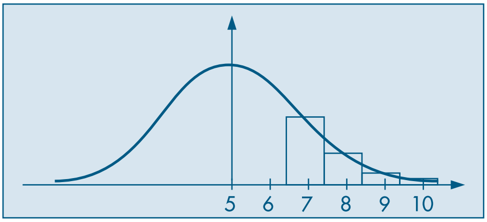
$$\mu = np = 10 \times \frac{1}{2} = 5$$

e variância

$$\sigma^2 = np(1 - p) = 10 \times \frac{1}{2} \times \frac{1}{2} = 2{,}5.$$

Veja a Figura 7.20.

Figura 7.20: Aproximação de $P(Y \geqslant 7)$ pela área sob a $N(5; 2{,}5)$.



Chamando X tal variável, com distribuição normal,

$$P(Y \geqslant 7) \simeq P(X \geqslant 6{,}5) = P\left(\frac{X-\mu}{\sigma} \geqslant \frac{6{,}5-\mu}{\sigma}\right)$$

$$P\left(Z \geqslant \frac{6{,}5-5}{\sqrt{2{,}5}}\right) = P(Z \geqslant 0{,}94) = 0{,}174,$$

onde Z é, como sempre, $N(0,\ 1)$. Utilizando a Tabela I, vemos que a probabilidade verdadeira é 0,172.

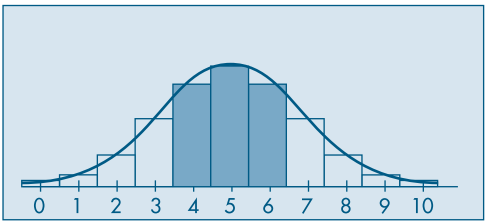
Vamos calcular agora $P(3 < Y \leqslant 6) = P(Y = 4) + P(Y = 5) + P(Y = 6)$. Vemos, através da Figura 7.21, que a aproximação a ser feita deve ser

$$P(3 < Y \leqslant 6) \simeq P(3{,}5 \leqslant X \leqslant 6{,}5) = P\left(\frac{3{,}5-5}{1{,}58} \leqslant Z \leqslant \frac{6{,}5-6}{1{,}58}\right)$$

$$= P(-0{,}94 \leqslant Z \leqslant 0{,}94) = 0{,}653,$$

ao passo que a probabilidade verdadeira é 0,656.

Figura 7.21: Aproximação de $P(3 < Y \leqslant 6)$.



A justificativa formal de tal aproximação é dada pelo chamado Teorema Limite Central, que será visto no Capítulo 10. A aproximação é boa quando $np > 5$ e $n(1-p) > 5$.

Problemas

13. A temperatura T de destilação do petróleo é crucial na determinação da qualidade final do produto. Suponha que T seja considerada uma v.a. com distribuição uniforme no intervalo (150, 300). Suponha que o custo para produzir um galão de petróleo

seja C_1 reais. Se o óleo for destilado a uma temperatura inferior a 200°, o produto obtido é vendido a C_2 reais; se a temperatura for superior a 200°, o produto é vendido a C_3 reais.

(a) Fazer o gráfico da f.d.p. de T.

(b) Qual o lucro médio por galão?

14. Se $X \sim N(10, 4)$, calcular:

(a) $P(8 < X < 10)$,

(b) $P(9 \leqslant X \leqslant 12)$,

(c) $P(X > 10)$,

(d) $P(X < 8 \text{ ou } X > 11)$.

15. Para $X \sim N(100, 100)$, calcule:

(a) $P(X < 115)$,

(b) $P(X \geqslant 80)$,

(c) $P(|X - 100| \leqslant 10)$,

(d) o valor a, tal que $P(100 - a \leqslant X \leqslant 100 + a) = 0{,}95$.

16. Para a v.a. $X \sim N(\mu, \sigma^2)$, encontre:

(a) $P(X \leqslant \mu + 2\sigma)$,

(b) $P(|X - \mu| \leqslant \sigma)$,

(c) o número a tal que $P(\mu - a\sigma \leqslant X \leqslant \mu + a\sigma) = 0{,}99$,

(d) o número b tal que $P(X > b) = 0{,}90$.

17. As alturas de 10.000 alunos de um colégio têm distribuição aproximadamente normal, com média 170 cm e desvio padrão 5 cm.

(a) Qual o número esperado de alunos com altura superior a 165 cm?

(b) Qual o intervalo simétrico em torno da média que conterá 75% das alturas dos alunos?

18. As vendas de determinado produto têm distribuição aproximadamente normal, com média 500 unidades e desvio padrão 50 unidades. Se a empresa decide fabricar 600 unidades no mês em estudo, qual é a probabilidade de que não possa atender a todos os pedidos desse mês, por estar com a produção esgotada?

19. Suponha que as amplitudes de vida de dois aparelhos elétricos, D_1 e D_2, tenham distribuições $N(42, 36)$ e $N(45, 9)$, respectivamente. Se os aparelhos são feitos para ser usados por um período de 45 horas, qual aparelho deve ser preferido? E se for por um período de 49 horas?

20. O diâmetro X de rolamentos esféricos produzidos por uma fábrica tem distribuição $N(0{,}6140; (0{,}0025)^2)$. O lucro T de cada rolamento depende de seu diâmetro. Assim,

$T = 0{,}10$, se o rolamento for bom $(0{,}610 < X < 0{,}618)$;

$T = 0{,}05$, se o rolamento for recuperável $(0{,}608 < X < 0{,}610)$ ou $(0{,}618 < X < 0{,}620)$;

$T = -0{,}10$, se o rolamento for defeituoso $(X < 0{,}608 \text{ ou } X > 0{,}620)$.

Calcule:

(a) as probabilidades de que os rolamentos sejam bons, recuperáveis e defeituosos.

(b) $E(T)$.

21. Suponha que um mecanismo eletrônico tenha um tempo de vida X (em 1.000 horas) que possa ser considerado uma v.a. contínua com f.d.p. $f(x) = e^{-x}$, $x > 0$. Suponha que o custo de fabricação de um item seja 2,00 reais e o preço de venda seja 5,00 reais. O fabricante garante total devolução se $X \leqslant 0{,}9$. Qual o lucro esperado por item?

22. Seja Y com distribuição binomial de parâmetros $n = 10$ e $p = 0{,}4$. Determine a aproximação normal para:
    (a) $P(3 < Y < 8)$, (b) $P(Y \geqslant 7)$, (c) $P(Y < 5)$.

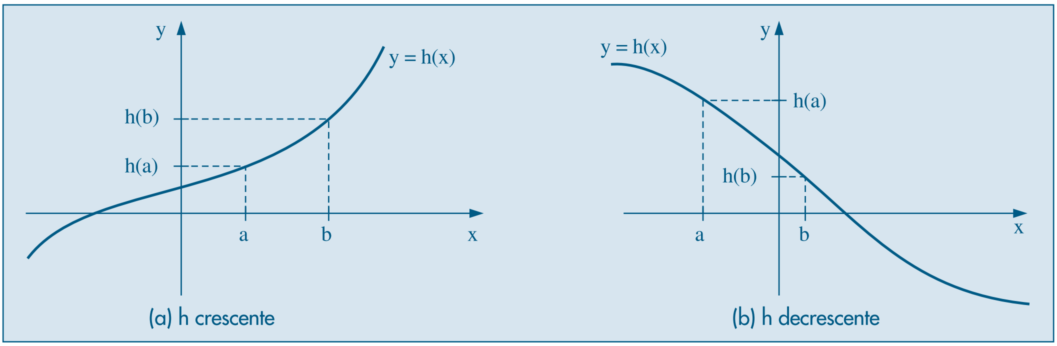
23. De um lote de produtos manufaturados, extraímos 100 itens ao acaso; se 10% dos itens do lote são defeituosos, calcule a probabilidade de 12 itens serem defeituosos. Use também a aproximação normal.

24. A confiabilidade de um mecanismo eletrônico é a probabilidade de que ele funcione sob as condições para as quais foi planejado. Uma amostra de 1.000 desses itens é escolhida ao acaso e os itens são testados, obtendo-se 30 defeituosos. Calcule a probabilidade de se obter pelo menos 30 itens defeituosos, supondo que a confiabilidade de cada item é 0,95.

7.6 Funções de Variáveis Contínuas

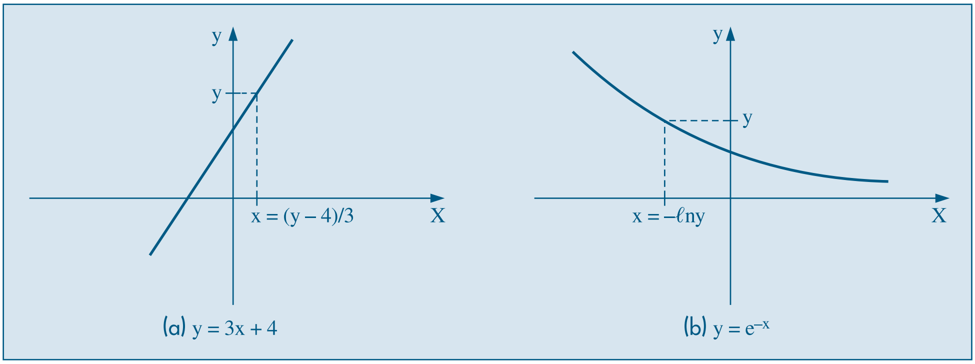
Vimos, no Capítulo 6, como obter a distribuição de uma v.a. $Y = h(X)$, se conhecermos a distribuição da v.a. discreta X. Vejamos, agora, o caso em que X é contínua. Suponhamos, primeiramente, que a função h seja estritamente monotônica, crescente ou decrescente. Neste caso, a inversa h^{-1} estará univocamente determinada e podemos obter $x = h^{-1}(y)$, para valores x e y das v.a. X e Y, respectivamente. Observando a Figura 7.22, vemos que, se a densidade de X, $f(x)$, digamos, for positiva no intervalo $a < x < b$, então a densidade de Y será positiva para $h(a) < y < h(b)$, se h for crescente, e para $h(b) < y < h(a)$, se h for decrescente.

Figura 7.22: Função de uma v.a.



Exemplo 7.11. Suponha X com a densidade do Exemplo 7.2 e considere $Y = 3X + 4$. Aqui, $y = h(x) = 3x + 4$, que é crescente (Figura 7.23 (a)).

Figura 7.23: Exemplos de funções de v.a. (a) Exemplo 7.11 (b) Exemplo 7.12.



Denotando a densidade de Y por $g(y)$, e como $f(x) > 0$ para $0 < x < 1$, $g(y) > 0$ para $4 < y < 7$.

Notemos que se podem obter probabilidades relativas a Y a partir da densidade de X. Por exemplo,

$$P(Y > 1) = P(3X + 4 > 1) = P(X > -1) = 1.$$

Vejamos como se pode obter $g(y)$. Denotemos por $G(y)$ a função de distribuição acumulada de Y. Da seção 7.3, sabemos que $G'(y) = g(y)$, para todo valor de y para o qual G for derivável. Então, temos

$$G(y) = P(Y \leqslant y) = P(3X + 4 \leqslant y) = P\left(X \leqslant \frac{y-4}{3}\right) = F\left(\frac{y-4}{3}\right),$$

onde estamos denotando por $F(\cdot)$ a função de distribuição acumulada de X. Usando a regra da cadeia para derivadas, temos

$$G'(y) = F'\left(\frac{y-4}{3}\right) \cdot \frac{1}{3} = \frac{1}{3} f\left(\frac{y-4}{3}\right),$$

do que decorre

$$g(y) = \begin{cases} \dfrac{2}{9}(y-4), & \text{se } 4 < y < 7 \\ 0, & \text{caso contrário.} \end{cases}$$

Exemplo 7.12. Suponha, agora, que X tenha densidade $f(x) = 3x^2/2$, $-1 < x < 1$ e que $Y = e^{-X}$. Segue-se que $h(x) = e^{-x}$ é uma função decrescente e $x = -\ell n(y)$ (Figura 7.23 (b)). Então,

$$\begin{aligned} G(y) &= P(Y \leqslant y) = P(e^{-X} \leqslant y) = P(X \geqslant -\ell n(y)) \\ &= 1 - P(X \leqslant -\ell n(y)) = 1 - F(-\ell n(y)), \end{aligned}$$

onde novamente F denota a f.d.a. de X. Derivando, obtemos a f.d.p. de Y,

$$g(y) = \frac{3}{2y}(\ell n(y))^2, \; e^{-1} < y < e.$$

O seguinte resultado generaliza esses dois exemplos.

Teorema 7.1. Se X for uma v.a. contínua, com densidade $f(x) > 0$, $a < x < b$, então $Y = h(X)$ tem densidade

$$g(y) = f(h^{-1}(y)) \left| \frac{dx}{dy} \right|, \tag{7.30}$$

supondo que h seja monotônica, derivável para todo x. Se h for crescente, $g(y) > 0$, $h(a) < y < h(b)$ e, se h for decrescente, $g(y) > 0$, $h(b) < y < h(a)$.

Prova. Basta notar que $G(y) = P(Y \leqslant y) = P(h(X) \leqslant y)$ e que essa probabilidade é igual a $P(X \leqslant h^{-1}(y)) = F(h^{-1}(y))$, se h for crescente, e igual a $1 - F(h^{-1}(y))$, se h for decrescente. Derivando $G(y)$ obtemos o resultado, notando que a derivada $(h^{-1}(y))' = dx/dy > 0$ se h for crescente, e negativa se h for decrescente.

Suponha, agora, que h não seja monotônica. Um caso de interesse que será usado mais tarde é $Y = h(X) = X^2$ (Figura 7.24). Temos

$$\begin{aligned} G(y) &= P(Y \leqslant y) = P(X^2 \leqslant y) = P(-\sqrt{y} \leqslant X \leqslant \sqrt{y}) \\ &= F(\sqrt{y}) - F(-\sqrt{y}), \end{aligned}$$

e derivando obtemos a densidade de Y,

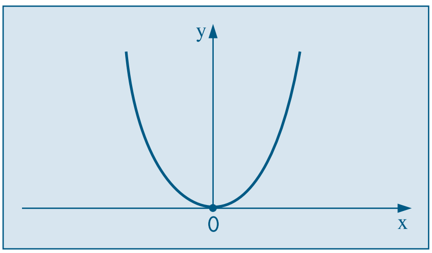
$$g(y) = \frac{1}{2\sqrt{y}}\,[\,f(\sqrt{y}) + f(-\sqrt{y})], \tag{7.31}$$

onde f é a densidade de X.

Se $f(x) = 1$, $0 < x < 1$ (X é uniforme no intervalo $[0, 1]$), então

$$g(y) = \frac{1}{2\sqrt{y}}, \; 0 < y < 1.$$

Figura 7.24: Ilustração de $Y = h(X) = X^2$.



Problemas

25. Considere a v.a. X do Problema 2 e $Y = X + 5$.
    (a) Calcule $P(Y \leqslant 5{,}5)$.
    (b) Obtenha a densidade de Y.
    (c) Obtenha a densidade de $Z = 2X$.

26. Suponha que a v.a. X tenha a densidade do Problema 8. Se $Y = 2X - 3/5$, obter a densidade de Y. Calcule $E(Y)$ e $\text{Var}(Y)$.

27. Suponha $X \sim \mathcal{U}[-1, 1]$. Calcule a densidade de $Y = X^2$ e de $W = |X|$.

7.7 Outros Modelos Importantes

Nesta seção vamos introduzir alguns modelos para v.a. contínuas que serão bastante utilizados na terceira parte deste livro. Juntamente com o modelo normal, esses modelos são úteis para as v.a. de interesse prático, que na maioria dos casos assumem valores positivos e tendem a ter distribuições assimétricas à direita.

7.7.1 A Distribuição Gama

Uma extensão do modelo exponencial é estudado a seguir.

Definição. A v.a. contínua X, assumindo valores positivos, tem uma distribuição gama com parâmetros $\alpha > 0$ e $\beta > 0$, se sua f.d.p. for dada por

$$f(x;\, \alpha,\, \beta) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\beta^{\alpha}}\, x^{\alpha-1}e^{-x/\beta}, & x > 0, \\ 0, & x < 0. \end{cases} \tag{7.32}$$

Em (7.32), $\Gamma(\alpha)$ é a *função gama*, importante em muitas áreas da Matemática, dada por

$$\Gamma(\alpha) = \int_0^{\infty} e^{-x}x^{\alpha-1}dx,\ \alpha > 0. \tag{7.33}$$

Não é difícil ver que $\Gamma(\alpha) = (\alpha - 1)\,\Gamma(\alpha - 1)$, se $\alpha = n$ for um inteiro positivo, $\Gamma(n) = (n - 1)!$ e que $\Gamma(1) = 1$, $\Gamma(1/2) = \sqrt{\pi}$. Veja o Problema 45.

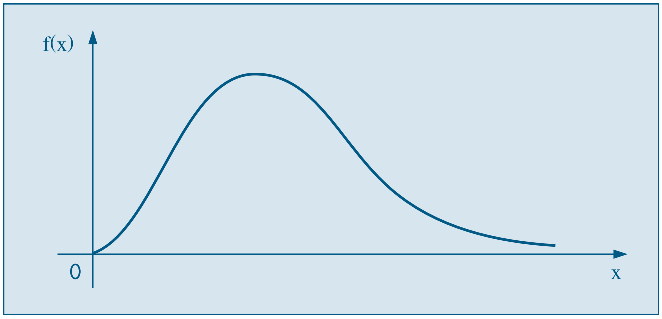
A Figura 7.25 ilustra a densidade (7.32) para $\alpha = 3$ e $\beta = 1$. Se $\alpha = 1$ obtemos a distribuição exponencial (7.26). Muitos casos de interesse têm α inteiro positivo.

Usaremos a notação

$$X \sim \text{Gama}(\alpha,\, \beta)$$

para designar uma v.a. com a distribuição dada por (7.32).

Figura 7.25: Gráfico da f.d.p. de uma distribuição gama, $\alpha = 3$, $\beta = 1$.



Pode-se demonstrar que:

$$E(X) = \alpha\beta,\ \mathrm{Var}(X) = \alpha\beta^2. \tag{7.34}$$

7.7.2 A Distribuição Qui-Quadrado

Um caso especial importante do modelo gama é obtido fazendo-se $\alpha = \nu/2$ e $\beta = 2$, com $\nu > 0$ inteiro.

Definição. Uma v.a. contínua Y, com valores positivos, tem uma distribuição qui-quadrado com ν graus de liberdade (denotada $\chi^2(\nu)$), se sua densidade for dada por

$$f(y;\ \nu) = \begin{cases} \dfrac{1}{\Gamma(\nu/2)2^{\nu/2}} y^{\nu/2\,-\,1} e^{-y/2}, & y > 0 \\ 0, & y < 0. \end{cases} \tag{7.35}$$

A Figura 7.26 ilustra os gráficos de (7.35) para $\nu = 1, 2, 3$. Segue-se de (7.34) que

$$E(Y) = \nu, \quad \mathrm{Var}(Y) = 2\nu. \tag{7.36}$$

A distribuição qui-quadrado tem muitas aplicações em Estatística e, como no caso da normal, existem tabelas para obter probabilidades. A Tabela IV, fornece os valores de y_0 tais que $P(Y > y_0) = p$, para alguns valores de p e de ν. Ver Figura 7.27.

Figura 7.26: Gráficos da distribuição qui-quadrado c²(n).

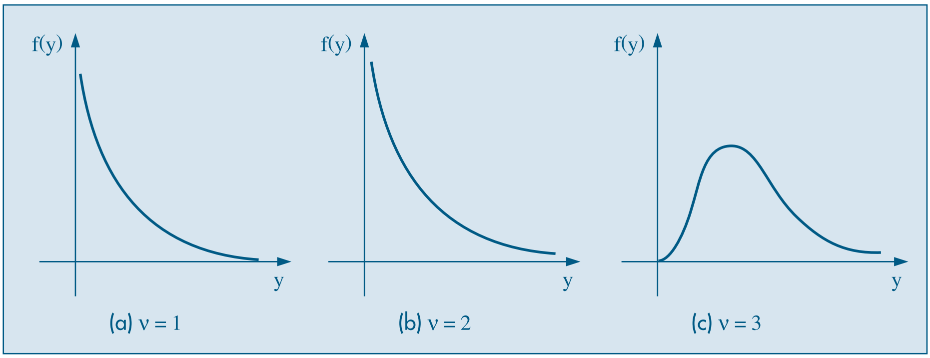
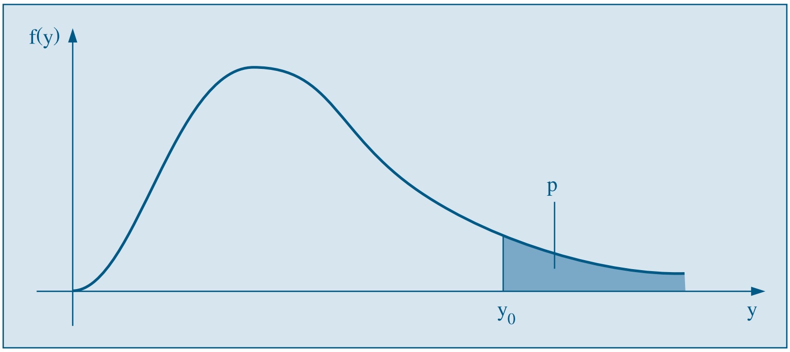


Figura 7.27: Valores tabelados da distribuição $\chi^2(\nu)$.



Exemplo 7.13. Usando a Tabela IV, para $\nu = 10$, observe que $P(Y > 2{,}558) = 0{,}99$, ao passo que $P(Y > 18{,}307) = 0{,}05$.

Para $\nu > 30$ podemos usar uma aproximação normal à distribuição qui-quadrado. Especificamente, temos o seguinte resultado: se Y tiver distribuição qui-quadrado com ν graus de liberdade, então a v.a.

$$Z = \sqrt{2Y} - \sqrt{2\nu - 1} \sim N(0,1).$$

Por exemplo, consultando a Tabela IV, temos que, se $\nu = 30$,

$$P(Y > 40{,}256) = 0{,}10,$$

enquanto que, usando a fórmula acima, temos que

$$z = \sqrt{2 \times 40{,}256} - \sqrt{59} = 1{,}292$$

e $P(Z > 1{,}292) = 0{,}099$, que resulta ser uma boa aproximação.

Exemplo 7.14. Considere $Z \sim N(0,1)$ e considere a v.a. $Y = Z^2$. De (7.31) temos que a densidade de Y é dada por

$$g(y) = \frac{1}{2\sqrt{y}}\,[\phi(\sqrt{y}) + \phi(-\sqrt{y})],\ y > 0,$$

onde por $\phi(z)$ indicamos a densidade da $N(0,1)$. Resulta

$$g(y) = \frac{1}{\sqrt{2\pi}}\, y^{-1/2}\, e^{-y/2},$$

e comparando com (7.35) vemos que $Y \sim \chi^2(1)$. Temos, aqui, um resultado importante: O quadrado de uma v.a. com distribuição normal padrão é uma v.a. com distribuição $\chi^2(1)$.

De um modo mais geral, uma v.a. x^2 (ν) *pode ser vista como a soma de* ν *normais padrões ao quadrado, independentes.*

7.7.3 A Distribuição *t* de Student

A distribuição t de Student é importante no que se refere a inferências sobre médias populacionais, tópico a ser tratado nos Capítulos 12 e 13. A obtenção da densidade está contida no teorema abaixo.

Teorema 7.1. Seja Z uma v.a. $N(0,1)$ e Y uma v.a. $\chi^2(\nu)$, com Z e Y independentes. Então, a v.a.

$$t = \frac{Z}{\sqrt{Y/\nu}}, \tag{7.37}$$

tem densidade dada por

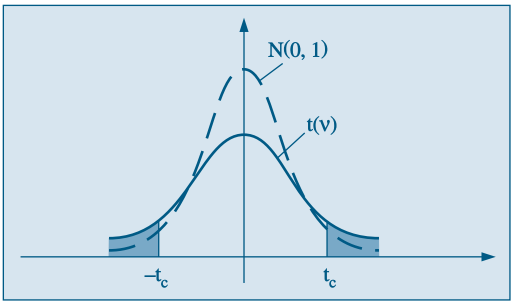
$$f(t;\, \nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}}\,(1 + t^2/\nu)^{-(\nu+1)/2},\ -\infty < t < \infty. \tag{7.38}$$

Diremos que tal variável tem uma *distribuição t de Student com ν graus de liberdade* e a indicaremos por $t(\nu)$. Pode-se provar que

$$E(t) = 0, \quad \text{Var}(t) = \frac{\nu}{\nu - 2},\ \nu > 2, \tag{7.39}$$

e verificar que o gráfico da densidade de t aproxima-se bastante de uma $N(0,1)$ quando ν é grande. Veja a Figura 7.28.

Figura 7.28: A distribuição t de Student e a distribuição normal padrão.



Como essa distribuição é bastante utilizada na prática, existem tabelas fornecendo probabilidades relativas a ela. A Tabela V fornece os valores de t_c tais que

$$P(-t_c < t(\nu) < t_c) = 1 - p, \tag{7.40}$$

para alguns valores de p e de ν.

O nome *Student* vem do pseudônimo usado pelo estatístico inglês W. S. Gosset, que introduziu essa distribuição no início do século passado.

Exemplo 7.15. Se $\nu = 6$, então, usando a Tabela V, $P(-1{,}943 < t(6) < 1{,}943) = 0{,}90$, ao passo que $P(t(6) > 2{,}447) = 0{,}025$. Observe que, nessa tabela, há uma linha com $\nu = \infty$, que corresponde a usar os valores da $N(0{,}1)$. Para $n > 120$ essa aproximação é muito boa.

7.7.4 A Distribuição *F* de Snedecor

Vamos considerar agora uma v.a. definida como o quociente de duas variáveis com distribuição qui-quadrado.

O seguinte teorema, que não será demonstrado, resume o que nos vai ser útil.

Teorema 7.2. Sejam U e V duas v.a. independentes, cada uma com distribuição qui-quadrado, com ν_1 e ν_2 graus de liberdade, respectivamente. Então, a v.a.

$$W = \frac{U/\nu_1}{V/\nu_2} \tag{7.41}$$

tem densidade dada por

$$g(w;\, \nu_1,\, \nu_2) = \frac{\Gamma((\nu_1 + \nu_2)/2)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)}\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \frac{w^{(\nu_1 - 2)/2}}{(1 + \nu_1 w/\nu_2)^{(\nu_1 + \nu_2)/2}},\ w > 0. \tag{7.42}$$

Diremos que W tem *distribuição F de Snedecor, com* ν_1 *e* ν_2 *graus de liberdade*, e usaremos a notação $W \sim F(\nu_1, \nu_2)$. Pode-se mostrar que

$$E(W) = \frac{\nu_2}{\nu_2 - 2} \quad \text{e} \quad \mathrm{Var}(W) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2\,(\nu_2 - 4)}. \tag{7.43}$$

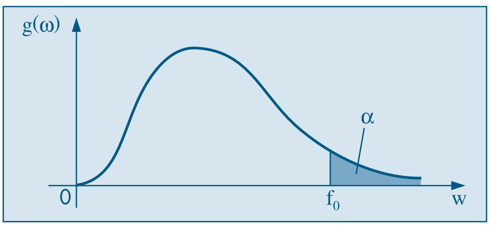
O gráfico típico de uma v.a. com distribuição F está na Figura 7.29. Na Tabela VI são dados os pontos f_0 tais que

$$P\{F(\nu_1, \nu_2) > f_0\} = \alpha,$$

para $\alpha = 0{,}05$, $\alpha = 0{,}025$ e alguns valores de ν_1 e ν_2. Para encontrar os valores inferiores, usa-se a identidade

$$F(\nu_1, \nu_2) = 1/F(\nu_2, \nu_1). \tag{7.44}$$

Figura 7.29: Gráfico de distribuição F.



Exemplo 7.16. Considere, por exemplo, $W \sim F(5, 7)$. Consultando a Tabela VI, $P(F > 3{,}97) = 0{,}05$ ou, então, $P(F \leqslant 3{,}97) = 0{,}95$. Digamos, agora, que desejamos encontrar o valor f_0 tal que $P(F < f_0) = 0{,}05$. Da igualdade (7.44) temos

$$0{,}05 = P\{F(5{,}7) < f_0\} = P\{1/F(7{,}5) < f_0\} = P\{F(7{,}5) > 1/f_0\},$$

e procurando na Tabela VI, para $F(7{,}5)$, obtemos $1/f_0 = 4{,}88$ e, portanto, $f_0 = 0{,}205$.

Na seção de Problemas e Complementos apresentamos algumas outras distribuições de interesse, como a log-normal, Pareto, Weibull e beta.

Na Tabela 7.2 mostramos os principais modelos para v.a. contínuas, incluindo: a densidade, o domínio dos valores, os parâmetros, a média e a variância.

Tabela 7.2: Modelos para variáveis contínuas.

Modelo	$f(x)$	Parâmetros	$E(X)$, $\mathrm{Var}(X)$
Uniforme	$1/(\beta-\alpha),\ \alpha < x < \beta$	α, β	$(\alpha+\beta)/2,\ (\beta-\alpha)^2/12$
Exponencial	$1/\beta\ e^{-t/\beta},\ t > 0$	β	$\beta,\ \beta^2$
Normal	$\frac{1}{\sigma\sqrt{2\pi}}\ \exp\left\{\left(\frac{x-\mu}{\sigma}\right)^2\right\}, -\infty < x < \infty$	μ, σ	$\mu,\ \sigma^2$
Gama	$\beta^{-\alpha}/\Gamma(\alpha)\ x^{\alpha-1}\ e^{-x/\beta},\ x > 0$	$\beta > 0,\ \alpha > 0$	$\alpha\beta,\ \alpha\beta^2$
Qui-quadrado	$\frac{2^{-\nu/2}}{\Gamma(\nu/2)}\ y^{\nu/2-1}e^{-y/2},\ y > 0$	ν	$\nu,\ 2\nu$
t-Student	$\frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi\nu}}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}, -\infty < t < \infty$	ν	$0,\ \nu/(\nu-2)$
F-Snedecor	$\frac{\Gamma\left(\frac{(\nu_1+\nu_2)}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}\frac{w^{\frac{\nu_1-2}{2}}}{\left(1+\frac{\nu_1 w}{\nu_2}\right)^{\frac{\nu_1+\nu_2}{2}}},\ w > 0.$	ν_1, ν_2	$\frac{\nu_2}{\nu_2-2},\ \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$

7.8 Quantis

No Capítulo 6 definimos o p-quantil $Q(p)$ como o valor da v.a. discreta X satisfazendo as duas desigualdades de (6.26).

No caso de uma v.a. contínua X, essa definição torna-se mais simples. Se $F(x)$ designar a f.d.a. de X, temos que as desigualdades em (6.26) ficam:

$$P(X \leqslant Q(p)) = F(Q(p)) \geqslant p \tag{7.45}$$

e

$$P(X \geqslant Q(p)) = 1 - P(X < Q(p)) = 1 - P(X \leqslant Q(p)) = 1 - F(Q(p)) \geqslant 1 - p. \tag{7.46}$$

Mas (7.46) pode ser reescrita como

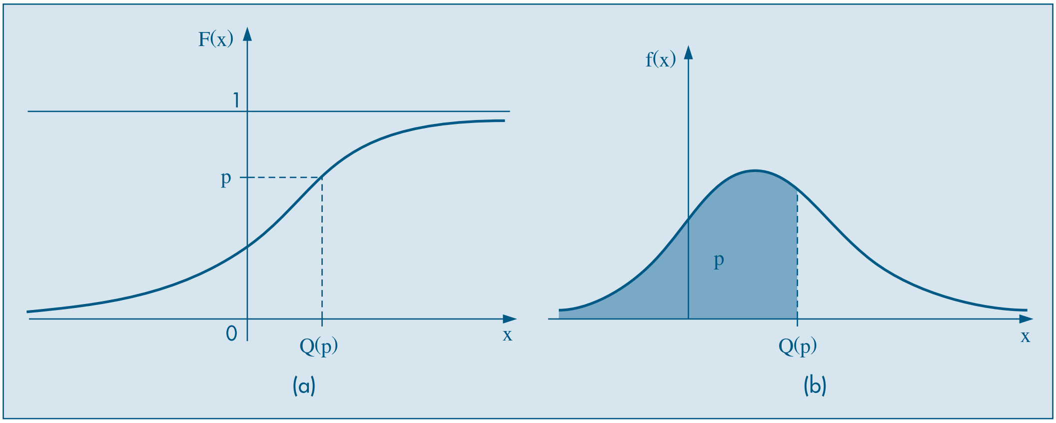
$$F(Q(p)) \leqslant p. \tag{7.47}$$

Portanto, de (7.45) e (7.47) chegamos à conclusão de que o p-quantil deve satisfazer

$$F(Q(p)) = p. \tag{7.48}$$

Graficamente, temos a situação ilustrada na Figura (7.30). Ou seja, para obter $Q(p)$, marcamos p no eixo das ordenadas, consideramos a reta horizontal pelo ponto $(0, p)$ até encontrar a curva de $F(x)$ e baixamos uma reta vertical até encontrar $Q(p)$ no eixo das abscissas. Analiticamente, temos de resolver a equação (7.48). Vejamos alguns exemplos.

Figura 7.30: Definição de Q(p) (a) f.d.a. (b) f.d.p.



Exemplo 7.17. Se $Z \sim N(0, 1)$, utilizando a Tabela III encontramos facilmente que

$Q(0, 5) = Q_2 = 0$,

$Q(0, 25) = Q_1 = -0{,}675$,

$Q(0, 30) = -0{,}52$,

$Q(0{,}75) = Q_3 = 0{,}675$.

Exemplo 7.18. Suponha que $Y \sim \text{Exp}(2)$. Se quisermos calcular a mediana, Q_2, teremos de resolver

$$\int_0^{Q_2} f(y)dy = 0{,}5,$$

Capítulo 8

Variáveis Aleatórias Multidimensionais

8.1 Distribuição Conjunta

Em muitas situações, ao descrevermos os resultados de um experimento, atribuímos a um mesmo ponto amostral os valores de duas ou mais variáveis aleatórias. Neste capítulo, iremos nos concentrar no estudo de um par de variáveis aleatórias, indicando que os conceitos e resultados apresentados estendem-se facilmente a um conjunto finito de variáveis aleatórias. Um tratamento mais completo é dado ao caso de variáveis discretas, nas seções 8.1 a 8.4.

Exemplo 8.1. Suponha que estamos interessados em estudar a composição de famílias com três crianças, quanto ao sexo. Definamos:

X = número de meninos,

$$Y = \begin{cases} 1, \text{ se o primeiro filho for homem} \\ 0, \text{ se o primeiro filho for mulher,} \end{cases}$$

Z = número de vezes em que houve variação do sexo entre um nascimento e outro, dentro da mesma família.

Com essas informações, e supondo que as possíveis composições tenham a mesma probabilidade, obtemos a Tabela 8.1, onde, por exemplo, o evento HMH indica que o primeiro filho é homem, o segundo, mulher e o terceiro, homem.

As distribuições de probabilidades das v.a. X, Y e Z podem ser obtidas dessa tabela e são dadas na Tabela 8.2.

Tabela 8.1: Composição de famílias com três crianças, quanto ao sexo.

Eventos	Probabilidade	X	Y	Z
HHH	1/8	3	1	0
HHM	1/8	2	1	1
HMH	1/8	2	1	2
MHH	1/8	2	0	1
HMM	1/8	1	1	1
MHM	1/8	1	0	2
MMH	1/8	1	0	1
MMM	1/8	0	0	0

Tabela 8.2: Distribuições de probabilidades unidimensionais.

(a)

x	0	1	2	3
$p(x)$	1/8	3/8	3/8	1/8

(b)

y	0	1
$p(y)$	1/2	1/2

(c)

z	0	1	2
$p(z)$	1/4	1/2	1/4

A Tabela 8.3 apresenta as probabilidades associadas aos pares de valores nas variáveis X e Y. Nessa tabela, $p(x, y) = P(X = x, Y = y)$ denota a probabilidade do evento $\{X = x \text{ e } Y = y\} = \{X = x\} \cap \{Y = y\}$. Essa tabela é denominada *distribuição conjunta* de X e Y.

Tabela 8.3: Distribuição bidimensional da v.a. (X, Y).

(x, y)	$p(x, y)$
(0, 0)	1/8
(1, 0)	2/8
(1, 1)	1/8
(2, 0)	1/8
(2, 1)	2/8
(3, 1)	1/8

A partir da Tabela 8.1 podemos formar também as distribuições conjuntas de X e Z, de Y e Z, bem como a distribuição conjunta de X, Y e Z, que está dada na Tabela 8.4.

Tabela 8.4: Distribuição conjunta das v.a. X, Y e Z.

(x, y, z)	$p(x, y, z)$
(0, 0, 0)	1/8
(1, 0, 1)	1/8
(1, 0, 2)	1/8
(1, 1, 1)	1/8
(2, 0, 1)	1/8
(2, 1, 1)	1/8
(2, 1, 2)	1/8
(3, 1, 0)	1/8

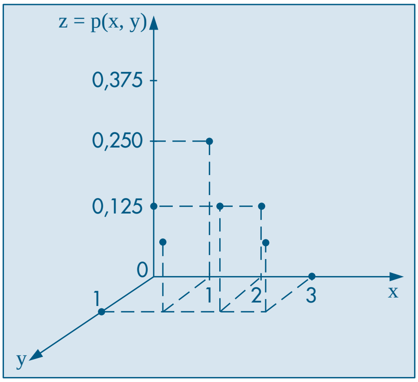
Aqui, $p(x, y, z) = P(X = x, Y = y, Z = z)$. Vamos nos fixar nas distribuições bidimensionais, isto é, nas distribuições conjuntas de duas variáveis. Nesse caso, uma maneira mais cômoda de representar a distribuição conjunta é por meio de tabelas de duplas entradas, como na Tabela 8.5, onde temos representada a mesma distribuição de X e Y, dada antes na Tabela 8.3.

Tabela 8.5: Distribuição conjunta de X e Y, como uma tabela de dupla entrada.

Y \ X	0	1	2	3	$p(y)$
0	1/8	2/8	1/8	0	1/2
1	0	1/8	2/8	1/8	1/2
$p(x)$	1/8	3/8	3/8	1/8	1

A representação gráfica de variáveis aleatórias bidimensionais (X, Y) exige gráficos com três eixos: um para a v.a. X, outro para a v.a. Y e um terceiro eixo z para a probabilidade conjunta $p(x, y)$. A Figura 8.1 representa a distribuição conjunta resumida na Tabela 8.5. A dificuldade em desenhar e interpretar tais gráficos nos leva, muitas vezes, a evitar o uso desse recurso tão valioso.

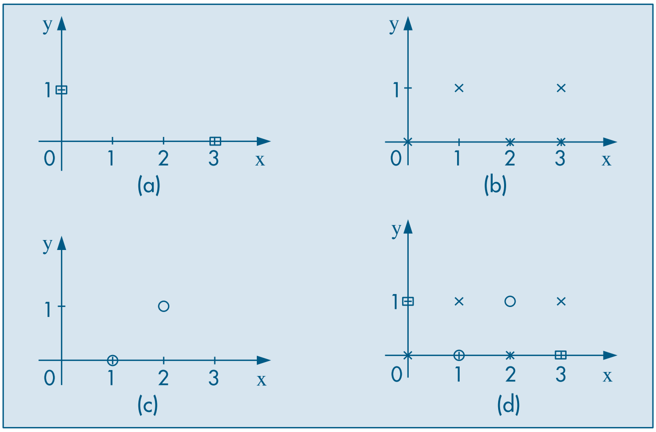
Figura 8.1: Representação gráfica da v.a. (X, Y) da Tabela 8.5.



Uma tentativa de representar distribuições de probabilidades discretas em duas dimensões é o gráfico de *curvas de níveis*. Esse é o mesmo recurso utilizado em mapas geográficos sobre relevos, indicando-se por meio de linhas as cotas (alturas) de mesma intensidade em uma região. Curvas de níveis podem ser usadas também em mapas meteorológicos, de marés etc.

Embora tais mapas sejam usados principalmente para variáveis contínuas, vamos exemplificar abaixo sua construção para os dados da Tabela 8.5. Notamos que existem valores apenas para as probabilidades 0, 1/8, 2/8 e 3/8, e cada um deles define um conjunto de pontos. Por exemplo, correspondendo à probabilidade 1/8 temos o conjunto de pontos (0, 0), (1, 1), (2, 0) e (3, 1). Na Figura 8.2 (b) representamos esses pontos, que corresponderiam à "curva de nível" para a cota 1/8. De modo análogo traçaríamos as demais curvas de níveis. A Figura 8.2 (e), reunindo todos os resultados, seria "equivalente" à Figura 8.1. Assim, os pontos representados por × formariam a curva de nível da cota 1/8; os pontos representados por ∘ formariam a curva de nível com cota (probabilidade) 2/8, e assim por diante. Esse recurso é mais bem visualizado para variáveis contínuas, como na Figura 8.17.

Figura 8.2: Curvas de níveis para a Tabela 8.5. (a) $p(x, y) = 0$ (b) $p(x, y) = 1/8$ (c) $p(x, y) = 2/8$ (d) todas as cotas



8.2 Distribuições Marginais e Condicionais

Da Tabela 8.5 podemos obter facilmente as distribuições de X e Y. A primeira e última colunas da tabela dão a distribuição de Y, $(y, p(y))$, enquanto a primeira e última linhas da tabela dão a distribuição de X, $(x, p(x))$. Essas distribuições são chamadas *distribuições marginais*.

Observamos, por exemplo, que

$$P(X = 1) = P(X = 1, Y = 0) + P(X = 1, Y = 1) = 2/8 + 1/8 = 3/8$$

e

$$P(Y = 0) = P(X = 0, Y = 0) + P(X = 1, Y = 0) + P(X = 2, Y = 0) + P(X = 3, Y = 0)$$
$$= 1/8 + 2/8 + 1/8 + 0 = 1/2.$$

Portanto, para obter as probabilidades marginais basta somar linhas e colunas.

Quando estudamos os aspectos descritivos das distribuições com mais de uma variável, vimos que, às vezes, é conveniente calcular proporções em relação a uma linha ou coluna, e não em relação ao total. Isso é equivalente aqui ao conceito de distribuição condicional. Por exemplo, qual seria a distribuição do número de meninos, sabendo-se que o primeiro filho é do sexo masculino? Ou seja, queremos calcular a probabilidade $P(X = x|Y = 1)$. Da definição de probabilidade condicional, obtemos

$$P(X = x|Y = 1) = \frac{P(X = x, Y = 1)}{P(Y = 1)} = p(x|Y = 1), \tag{8.1}$$

para $x = 0, 1, 2, 3$. Pela Tabela 8.5 obtemos, por exemplo,

$$p(2|Y=1) = P(X=2|Y=1) = \frac{P(X=2,\ Y=1)}{P(Y=1)} = \frac{2/8}{1/2} = 1/2.$$

Do mesmo modo, obtemos as demais probabilidades, e a *distribuição condicional* de X, *dado que* $Y = 1$, está na Tabela 8.6.

Tabela 8.6: Distribuição condicional de X, dado que $Y = 1$.

x	1	2	3
$p(x\|Y=1)$	1/4	1/2	1/4

Observe que $\sum_x p(x|Y=1) = p(0|Y=1) + \ldots + p(3|Y=1) = 1$.

Do mesmo modo, podemos obter a distribuição condicional de Y, dado que $X = 2$, que está na Tabela 8.7.

Tabela 8.7: Distribuição condicional de Y, dado que $X = 2$.

y	0	1
$p(y\|X=2)$	1/3	2/3

Podemos generalizar o que foi dito acima para duas v.a. X e Y quaisquer, assumindo os valores $x_1, x_2, \ldots, x_n$ e $y_1, y_2, \ldots, y_m$, respectivamente.

Definição. Seja x_i, um valor de X, tal que $P(X = x_i) = p(x_i) > 0$. A probabilidade

$$P(Y = y_j|X = x_i) = \frac{P(X = x_i,\ Y = y_j)}{P(X = x_i)}, \quad j = 1, \ldots, m, \tag{8.2}$$

é denominada *probabilidade condicional* de $Y = y_j$, *dado que* $X = x_i$.

Como observamos acima, para x_i fixado, os pares $(y_j, P(Y = y_j|X = x_i))$, $j = 1, \ldots, m$, definem a distribuição condicional de Y, dado que $X = x_i$, pois

$$\sum_{j=1}^{m} P(Y = y_j|X = x_i) = \sum_{j=1}^{m} \frac{P(Y = y_j,\ X = x_i)}{P(X = x_i)} = \frac{P(X = x_i)}{P(X = x_i)} = 1.$$

Considere a distribuição condicional de X, dado que $Y = 1$, da Tabela 8.6. Podemos calcular a média dessa distribuição, a saber

$$E(X|Y=1) = 1 \times \frac{1}{4} + 2 \times \frac{1}{2} + 3 \times \frac{1}{4} = 2.$$

Observe que $E(X) = 1{,}5$, ao passo que $E(X|Y=1) = 2$.

De modo geral temos a seguinte definição.

Definição. A esperança condicional de X, dado que $Y = y_j$, é definida por

$$E(X|Y = y_j) = \sum_{i=1}^{n} x_i P(X = x_i|Y = y_j).$$

Uma definição análoga vale para $E(Y|X = x_i)$.

Exemplo 8.2. Para a distribuição condicional de Y, dado que $X = 2$, da Tabela 8.7, temos

$$E(Y|X = 2) = 0 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{2}{3}.$$

Exemplo 8.3. Considere, agora, a distribuição conjunta das variáveis Y e Z, definidas no Exemplo 8.1. Da Tabela 8.1 obtemos a Tabela 8.8. Aqui, observamos que

$$P(Z = z|Y = y) = \frac{P(Z = z, Y = y)}{P(Y = y)} = P(Z = z)$$

para quaisquer $z = 0, 1, 2$ e $y = 0, 1$. O que significa dizer que

$$P(Z = z, Y = y) = P(Z = z)\, P(Y = y),$$

isto é, a probabilidade de cada casela é igual ao produto das respectivas probabilidades marginais. Por exemplo,

$$P(Z = 1, Y = 1) = \frac{2}{8} = \frac{2}{4} \times \frac{1}{2} = P(Z = 1)P(Y = 1).$$

Tabela 8.8: Distribuição conjunta de Y e Z.

Y \ Z	0	1	2	$p(y)$
0	1/8	2/8	1/8	1/2
1	1/8	2/8	1/8	1/2
$p(z)$	1/4	2/4	1/4	1

Também é verdade que

$$P(Y = y|Z = z) = P(Y = y)$$

para todos os valores de y e z. Dizemos que Y e Z são independentes.

Definição. As variáveis aleatórias X e Y, assumindo os valores $x_1, x_2, \ldots$ e $y_1, y_2, \ldots$, respectivamente, são independentes se, e somente se, para todo par de valores (x_i, y_j) de X e Y, tivermos que

$$P(X = x_i, Y = y_j) = P(X = x_i)\, P(Y = y_j). \tag{8.3}$$

Basta que (8.3) não se verifique para *um* par (x_i, y_j), para que X e Y *não* sejam independentes. Nesse caso diremos que X e Y são *dependentes*.

Essa definição pode ser estendida para mais de duas variáveis aleatórias.

Problemas

1. Lançam-se, simultaneamente, uma moeda e um dado.
   (a) Determine o espaço amostral correspondente a esse experimento.
   (b) Obtenha a tabela da distribuição conjunta, considerando X o número de caras no lançamento da moeda e Y o número da face do dado.
   (c) Verifique se X e Y são independentes.
   (d) Calcule:
      1. $P(X = 1)$
      2. $P(X \leqslant 1)$
      3. $P(X < 1)$
      4. $P(X = 2, Y = 3)$
      5. $P(X \geqslant 0, Y \leqslant 4)$
      6. $P(X = 0, Y \geqslant 1)$

2. A tabela abaixo dá a distribuição conjunta de X e Y.
   (a) Determine as distribuições marginais de X e Y.
   (b) Obtenha as esperanças e variâncias de X e Y.
   (c) Verifique se X e Y são independentes.
   (d) Calcule $P(X = 1|Y = 0)$ e $P(Y = 2|X = 3)$.
   (e) Calcule $P(X \leqslant 2)$ e $P(X = 2, Y \leqslant 1)$.

$Y \backslash X$	1	2	3
0	0,1	0,1	0,1
1	0,2	0	0,3
2	0	0,1	0,1

3. Considere a distribuição conjunta de X e Y, parcialmente conhecida, dada na tabela abaixo.
   (a) Complete a tabela, considerando X e Y independentes.
   (b) Calcule as médias e variâncias de X e Y.
   (c) Obtenha as distribuições condicionais de X, dado que $Y = 0$, e de Y, dado que $X = 1$.

$Y \backslash X$	–1	0	1	$P(Y = y)$
–1	1/12			
0				1/3
1	1/4		1/4	
$P(X = x)$				1

8.3 Funções de Variáveis Aleatórias

Retomemos a Tabela 8.5, que dá a distribuição conjunta das variáveis aleatórias X e Y. A partir dela, podemos considerar, por exemplo, a v.a. $X + Y$, ou a v.a. XY. A soma $X + Y$ é definida naturalmente: a cada resultado do experimento, ela associa a soma dos valores de X e Y, isto é,

$$(X + Y)(\omega) = X(\omega) + Y(\omega). \tag{8.4}$$

Do mesmo modo,

$$(XY)(\omega) = X(\omega)\, Y(\omega). \tag{8.5}$$

Podemos, então, construir a Tabela 8.9.

Tabela 8.9: Funções de variáveis aleatórias.

(x_i, y_j)	$X + Y$	XY	$p(x_i, y_j)$
(0, 0)	0	0	1/8
(0, 1)	1	0	0
(1, 0)	1	0	2/8
(1, 1)	2	1	1/8
(2, 0)	2	0	1/8
(2, 1)	3	2	2/8
(3, 0)	3	0	0
(3, 1)	4	3	1/8

A partir dessa tabela, obtemos as distribuições de $X + Y$ e XY, ilustradas nas Tabelas 8.10 e 8.11.

Tabela 8.10: Distribuição de $X + Y$.

$x + y$	0	1	2	3	4
$p(x + y)$	1/8	2/8	2/8	2/8	1/8

Tabela 8.11: Distribuição de XY.

xy	0	1	2	3
$p(xy)$	4/8	1/8	2/8	1/8

Vimos, no Capítulo 6, como calcular a esperança de uma v.a. Para as v.a X e Y da Tabela 8.5, temos:

$$E(X) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{12}{8} = 1{,}5,$$

$$E(Y) = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = 0{,}5.$$

Da Tabela 8.10, obtemos

$$E(X + Y) = 0 \times \frac{1}{8} + 1 \times \frac{2}{8} + 2 \times \frac{2}{8} + 3 \times \frac{2}{8} + 4 \times \frac{1}{8} = \frac{16}{8} = 2.$$

Notamos que $E(X + Y) = E(X) + E(Y)$. Poderia ser uma simples coincidência, mas essa relação é de fato verdadeira.

Teorema 8.1. Se X for uma v.a. com valores $x_1, \ldots, x_n$ e probabilidades $p(x_1), \ldots, p(x_n)$, Y for uma v.a. com valores $y_1, \ldots, y_m$ e probabilidades $p(y_1), \ldots, p(y_m)$, e se $p(x_i, y_j) = P(X = x_i, Y = y_j)$, $i = 1, \ldots, n$, $j = 1, \ldots, m$, então

$$E(X + Y) = E(X) + E(Y). \tag{8.6}$$

Prova. Observando a Tabela 8.9, podemos escrever

$$E(X+Y) = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i + y_j)p(x_i, y_j)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m}x_i p(x_i, y_j) + \sum_{i=1}^{n}\sum_{j=1}^{m}y_j p(x_i, y_j). \tag{8.7}$$

Mas, para um i fixo, $\sum_{j=1}^{m} p(x_i, y_j) = p(x_i)$, e para um j fixo, $\sum_{i=1}^{n} p(x_i, y_j) = p(y_j)$, logo, podemos escrever

$$E(X) = \sum_{i=1}^{n} x_i p(x_i) = \sum_{i=1}^{n} x_i \sum_{j=1}^{m} p(x_i, y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i p(x_i, y_j)$$

e

$$E(Y) = \sum_{j=1}^{m} y_j p(y_j) = \sum_{j=1}^{m} y_j \sum_{i=1}^{n} p(x_i, y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} y_j p(x_i, y_j).$$

Comparando essas duas últimas relações com (8.7), obtemos a relação (8.6).

Do que foi visto acima, podemos concluir que, se X e Y são duas v.a. nas condições do Teorema 8.1, e se $g(X, Y)$ for uma função de X e Y, então

$$E[g(X, Y)] = \sum_{i=1}^{n}\sum_{j=1}^{m} g(x_i, y_j)p(x_i, y_j). \tag{8.8}$$

Exemplo 8.4. Da Tabela 8.9 temos

$$E(XY) = 0 \times \frac{1}{8} + 0 \times 0 + 0 \times \frac{2}{8} + 1 \times \frac{1}{8} + 2 \times \frac{2}{8} + 0 \times 0$$
$$+ 3 \times \frac{1}{8} = \frac{8}{8} = 1.$$

É claro que o mesmo valor pode ser obtido da Tabela 8.11, isto é, se $Z = XY$ e $p(z) = p(xy)$, então

$$E(Z) = E(XY) = 0 \times \frac{4}{8} + 1 \times \frac{1}{8} + 2 \times \frac{2}{8} + 3 \times \frac{1}{8} = 1.$$

Observamos que, neste caso,

$$E(Z) = E(XY) = 1 \neq E(X)E(Y) = (1{,}5)\,(0{,}5) = 0{,}75,$$

ou seja, de modo geral, a esperança de um produto de duas v.a. não é igual ao produto das esperanças das v.a. No entanto, existem situações em que essa propriedade se verifica. O teorema seguinte apresenta uma dessas situações.

Teorema 8.2. Se X e Y são variáveis aleatórias independentes, então

$$E(XY) = E(X)\,E(Y). \tag{8.9}$$

Prova. Nas condições do Teorema 8.1, usando (8.8) e (8.3),

$$E(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p(x_i, y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j p(x_i) p(y_j),$$

logo,

$$E(XY) = \sum_{i=1}^{n} x_i p(x_i) \sum_{j=1}^{m} y_j p(y_j) = E(X)E(Y).$$

A recíproca do Teorema 8.2 não é verdadeira, isto é, (8.9) pode ser válida e X e Y serem dependentes. Veja o Exemplo 8.7 abaixo.

Observações. (i) Se tivermos um número finito de v.a. $X_1, \ldots, X_n$, então (8.6) toma a forma

$$E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n). \tag{8.10}$$

(ii) Se $X_1, \ldots, X_n$ forem v.a. independentes, então

$$E(X_1 X_2 \ldots X_n) = E(X_1)\,E(X_2) \ldots E(X_n). \tag{8.11}$$

Exemplo 8.5. Nas seções 6.6.2 e 6.6.3 definimos a v.a. de Bernoulli e a v.a. binomial. Seja X o número de sucessos em n provas de Bernoulli. Definamos

$$X_i = \begin{cases} 1, \text{ se no } i\text{-ésimo ensaio ocorreu sucesso} \\ 0, \text{ se no } i\text{-ésimo ensaio ocorreu fracasso,} \end{cases}$$

$i = 1, 2, \ldots, n$. Então, segue-se que

$$X = X_1 + X_2 + \ldots + X_n,$$

e $X_1, \ldots, X_n$ são independentes. Se $p = P(\text{sucesso})$, então

$$E(X_i) = 1 \times p + 0 \times (1 - p) = p,\ i = 1, \ldots, n$$

e, por (8.10),

$$E(X) = E(X_1) + \ldots + E(X_n) = np,$$

o que demonstra a relação (6.16). A relação (6.17) será demonstrada na seção seguinte.

Problemas

4. No Problema 2, obtenha as distribuições de $X + Y$ e de XY. Calcule $E(X + Y)$, $E(XY)$, $\text{Var}(X+Y)$, $\text{Var}(XY)$.

5. (a) No Problema 3, calcule $E(X+Y)$ e $\text{Var}(X+Y)$.
   (b) Se $Z = aX + bY$, calcule a e b de modo que $E(Z) = 10$ e $\text{Var}(Z) = 600$.

6. Dois tetraedros (dados com quatro faces) com as faces numeradas de um a quatro são lançados e os números das faces voltadas para baixo são anotados. Sejam as v.a.:
   X: maior dos números observados;
   Y: menor dos números observados;
   $Z = X + Y$.
   (a) Construa a tabela da distribuição conjunta de X e Y.
   (b) Determine as médias e as variâncias de X, Y e Z.

7. Numa urna têm-se cinco tiras de papel, numeradas 1, 3, 5, 5, 7. Uma tira é sorteada e recolocada na urna; então, uma segunda tira é sorteada. Sejam X_1 e X_2 o primeiro e o segundo números sorteados.
   (a) Determine a distribuição conjunta de X_1 e X_2.
   (b) Obtenha as distribuições marginais de X_1 e X_2. Elas são independentes?
   (c) Encontre a média e a variância de X_1, X_2 e $\overline{X} = (X_1 + X_2)/2$.
   (d) Como seriam as respostas anteriores se a primeira tira de papel não fosse devolvida à urna antes da segunda extração?

8. Numa urna têm-se cinco bolas marcadas com os seguintes números: -1, 0, 0, 0, 1. Retiram-se três bolas, simultaneamente; X indica a soma dos números extraídos e Y o maior valor da trinca. Calcule:
   (a) Função de probabilidade de (X, Y).
   (b) $E(X)$ e $\text{Var}(X)$.
   (c) $\text{Var}(X+Y)$.

9. Dada a distribuição conjunta de X e Y abaixo, determine a média e a variância de:
   (a) $X + Y$.
   (b) XY.

Y \ X	1	2	3
1	5/27	1/27	3/27
2	4/27	3/27	4/27
3	2/27	3/27	2/27

10. Suponha que X e Y tenham a seguinte distribuição conjunta:

Y \ X	1	2	3
1	0,1	0,1	0,0
2	0,1	0,2	0,3
3	0,1	0,1	0,0

(a) Determine a f.p. de $X+Y$ e, a partir dela, calcule $E(X+Y)$. Pode-se obter a mesma resposta de outra maneira?
(b) Determine a f.p. de XY e, em seguida, calcule $E(XY)$.
(c) Mostre que, embora $E(XY) = E(X)\,E(Y)$, X e Y não são independentes.

8.4 Covariância entre Duas Variáveis Aleatórias

Vamos introduzir agora uma medida da relação linear entre duas variáveis aleatórias.

Definição. Se X e Y são duas v.a., a covariância entre elas é definida por

$$\mathrm{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))], \tag{8.12}$$

ou seja, o valor médio do produto dos desvios de X e Y em relação às suas respectivas médias.

Suponha que X assuma os valores $x_1, \ldots, x_n$, e Y os valores $y_1, \ldots, y_m$, e que $P(X = x_i, Y = y_j) = p(x_i, y_j)$. Então, (8.12) pode ser escrita

$$\mathrm{Cov}(X, Y) = \sum_{i=1}^{n}\sum_{j=1}^{m}[x_i - E(X)][y_j - E(Y)]\,p(x_i, y_j). \tag{8.13}$$

A fórmula (8.12) pode ser escrita de uma forma mais simples. Note que

$$\begin{aligned}\mathrm{Cov}(X, Y) &= E[XY - XE(Y) - YE(X) + E(X)E(Y)]\\ &= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y),\end{aligned}$$

ou seja,

$$\mathrm{Cov}(X, Y) = E(XY) - E(X)\,E(Y). \tag{8.14}$$

Exemplo 8.6. Para as v.a. X e Y do Exemplo 8.1 (veja a Tabela 8.5), obtemos

$$E(X) = 1{,}5,\ E(Y) = 0{,}5,\ E(XY) = 1{,}0,$$

de modo que

$$\mathrm{Cov}(X, Y) = 1{,}0 - (1{,}5)\,(0{,}5) = 0{,}25.$$

Definição. Quando $\mathrm{Cov}(X, Y) = 0$, dizemos que as variáveis aleatórias X e Y são *não correlacionadas*.

Exemplo 8.7. Consideremos a distribuição conjunta de X e Y dada pela Tabela 8.12.

Tabela 8.12: Distribuição conjunta para o Exemplo 8.7.

$Y \backslash X$	0	1	2	$p(y)$
1	3/20	3/20	2/20	8/20
2	1/20	1/20	2/20	4/20
3	4/20	1/20	3/20	8/20
$p(x)$	8/20	5/20	7/20	1,00

Temos que:

$$E(X) = 0 \times \frac{8}{20} + 1 \times \frac{5}{20} + 2 \times \frac{7}{20} = 0{,}95,$$

$$E(Y) = 1 \times \frac{8}{20} + 2 \times \frac{4}{20} + 3 \times \frac{8}{20} = 2{,}00,$$

$$E(XY) = 0 \times \frac{3}{20} + 1 \times \frac{3}{20} + 2 \times \frac{2}{20} + 0 \times \frac{1}{20} + 2 \times \frac{1}{20}$$
$$+ 4 \times \frac{2}{20} + 0 \times \frac{4}{20} + 3 \times \frac{1}{20} + 6 \times \frac{3}{20} = 1{,}90,$$

do que obtemos

$$\text{Cov}(X, Y) = 1{,}90 - (0,\ 95)(2{,}00) = 0.$$

Portanto, as v.a. X e Y desse exemplo são não-correlacionadas.

Exemplo 8.8. Retomemos o Exemplo 8.3, para o qual vimos que Y e Z são independentes. É fácil ver que $E(Z) = 1$ e $E(Y) = 1/2$. Da Tabela 8.8 obtemos que $E(YZ) = 1/2$, do que decorre que a covariância entre Y e Z é zero.

De modo geral, se X e Y forem independentes, então (8.9) é válida, logo, por (8.14) temos que $\text{Cov}(X, Y) = 0$.

Vamos destacar esse fato por meio da

Proposição 8.1. Se X e Y são duas variáveis aleatórias independentes, então $\text{Cov}(X, Y) = 0$.

Em outras palavras, se X e Y forem independentes, então elas serão não-correlacionadas. A recíproca não é verdadeira, isto é, se tivermos $\text{Cov}(X, Y) = 0$, isso não implica que X e Y sejam independentes. De fato, para as v.a. do Exemplo 8.7, a covariância entre X e Y é zero, mas X e Y não são independentes, como podemos facilmente verificar.

Podemos agora demonstrar o

Teorema 8.3. *(a)* Para duas v.a. X e Y quaisquer, temos

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y); \quad (8.15)$$

(b) se X e Y forem independentes, então

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y). \tag{8.16}$$

Prova.

(a) $\mathrm{Var}(X + Y) = E[(X + Y) - E(X + Y)]^2$

$= E[X - E(X) + Y - E(Y)]^2 = \mathrm{E}[X - E(X)]^2 + E[Y - E(Y)]^2 + 2E[(X - E(X))(Y - E(Y))]$,

e da definição de covariância, obtemos (8.15).

(b) A relação (8.16) segue imediatamente da Proposição 8.1.

As relações (8.15) e (8.16) podem ser generalizadas para mais de duas variáveis. Em particular, se $X_1, \ldots, X_n$ são v.a. independentes, então

$$\mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n). \tag{8.17}$$

Exemplo 8.5. (continuação) Temos que

$$\mathrm{Var}(X_i) = p(1 - p), \text{ para todo } i = 1, \ldots, n,$$

logo

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n) = np(1 - p),$$

o que demonstra a relação (6.17).

Vamos introduzir agora uma medida que não depende das unidades de medida de X e Y. O análogo descritivo para dois conjuntos de dados foi introduzido na seção 4.5.

Definição. O coeficiente de correlação entre X e Y é definido por

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)} \tag{8.18}$$

Exemplo 8.9. Para X e Y do Exemplo 8.7, a covariância entre X e Y é zero, logo $\rho(X, Y) = 0$. Para X e Y do Exemplo 8.6, temos que $\mathrm{Cov}(X, Y) = 0{,}25$. Verifique que $\mathrm{Var}(X) = 0{,}75$, $\mathrm{Var}(Y) = 0{,}25$, logo

$$\rho(X, Y) = \frac{0{,}25}{\sqrt{(0{,}75)(0{,}25)}} = 0{,}58.$$

O seguinte resultado será demonstrado no Problema 48.

Teorema 8.4. O coeficiente de correlação entre X e Y satisfaz a desigualdade

$$-1 \leqslant \rho(X, Y) \leqslant 1.$$

O coeficiente de correlação é uma medida da relação *linear* entre X e Y. Quando $\rho(X, Y) = \pm 1$, existe uma correlação perfeita entre X e Y, pois $Y = aX + b$. Se $\rho(X, Y) = 1$, $a > 0$, e se $\rho(X, Y) = -1$, $a < 0$. O grau de associação linear entre X e Y varia à medida que $\rho(X, Y)$ varia entre -1 e $+1$.

As seguintes propriedades podem ser provadas facilmente (ver Problema 38). Se a e b são constantes, então:

$$\rho(X + a, Y + b) = \rho(X, Y), \tag{8.19}$$

$$\rho(aX, bY) = \frac{ab}{|ab|}\,\rho(X, Y). \tag{8.20}$$

Ou seja, se $ab > 0$, $\rho(aX, bY) = \rho(X, Y)$ e se $ab < 0$, $\rho(aX, bY) = -\rho(X, Y)$.

Exemplo 8.10. Ainda usando o enunciado do Exemplo 8.1, defina a v.a. W como sendo o "número de meninas". A distribuição conjunta de X e W está na Tabela 8.13.

Tabela 8.13: Distribuição conjunta de X e W para o Exemplo 8.10.

W \ X	0	1	2	3	$p(w)$
0	0	0	0	1/8	1/8
1	0	0	3/8	0	3/8
2	0	3/8	0	0	3/8
3	1/8	0	0	0	1/8
$p(x)$	1/8	3/8	3/8	1/8	1

É fácil ver que

$$E(X) = E(W) = 1{,}5,$$
$$\mathrm{Var}(X) = \mathrm{Var}(W) = 0{,}75,$$
$$E(XW) = 1{,}5,$$

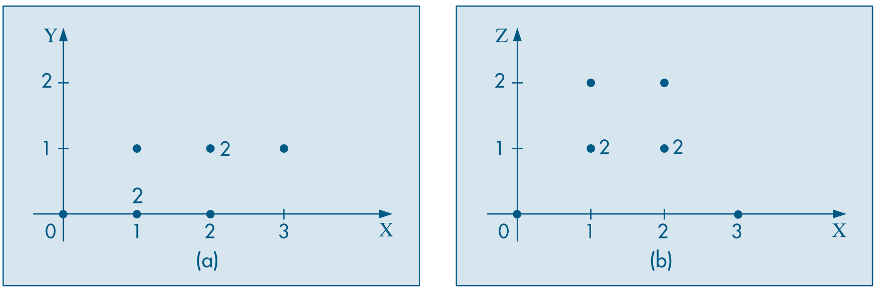
do que segue que $\mathrm{Cov}(X, W) = -0{,}75$ e portanto $\rho(X, W) = -1$. Esse é um resultado esperado, pois sabemos que $X = 3 - W$.

Para se analisar a possível correlação entre duas v.a. X e Y é conveniente usar os chamados *diagramas de dispersão*, que consistem no gráfico dos pares de valores de X e Y.

Exemplo 8.11. Na Figura 8.3 temos os diagramas de dispersão para as v.a. X e Y e X e Z, do Exemplo 8.1.

Figura 8.3: Diagramas de dispersão para as v.a. do Exemplo 8.1.

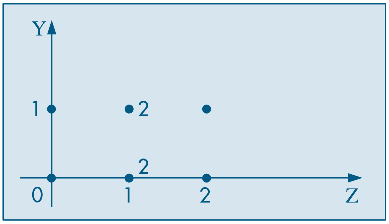
(a) X e Y (b) X e Z



Na Figura 8.3(a), ao lado dos pontos (1, 0) e (2, 1), colocamos o número 2, para mostrar que esses pares têm probabilidades 2/8, ao passo que os demais têm probabilidades 1/8.

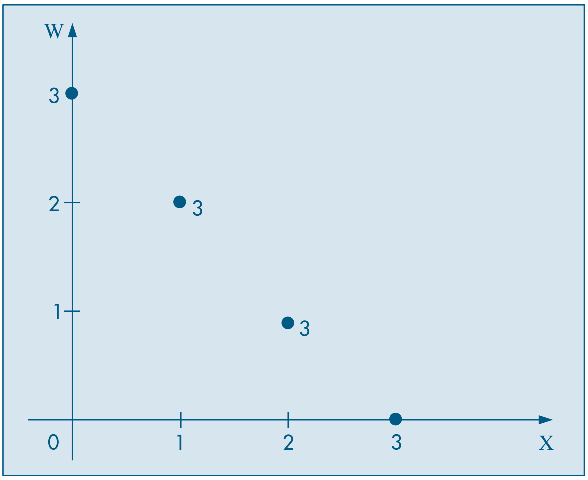
Exemplo 8.12. O diagrama de dispersão das v.a. Y e Z do Exemplo 8.2 está ilustrado na Figura 8.4. Lembremos que, nesse caso, Y e Z são independentes.

Figura 8.4: Diagrama de dispersão para as v.a. Y e Z do Exemplo 8.2.



Exemplo 8.13. Na Figura 8.5 temos o diagrama de dispersão das variáveis X e W do Exemplo 8.10. Observe que, nesse caso, existe uma relação linear perfeita entre as duas variáveis.

Figura 8.5: Diagrama de dispersão para as v.a. X e W do Exemplo 8.10.



Problemas

11. Para as v.a. X e Y do Problema 2 e usando os resultados do Problema 4, calcule $\text{Cov}(X, Y)$ e $\rho(X, Y)$.

12. Considere a situação do Problema 10 do Capítulo 6.
    (a) Obtenha as distribuições de $X + Y$ e $|X - Y|$.
    (b) Calcule $E(XY)$, $E(X/Y)$ e $E(X + Y)$.
    (c) Verifique se X e Y são independentes.

(*d*) Verifique se $E(XY) = E(X)\,E(Y)$. O que você pode concluir?
(*e*) Verifique se $E(X/Y) = E(X)/E(Y)$.
(*f*) Calcule $\text{Var}(X+Y)$. É verdade que $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$?

13. Sejam X e Y com a distribuição conjunta da tabela abaixo. Mostre que $\text{Cov}(X,Y) = 0$, mas X e Y não são independentes.

Y \ X	–1	0	1
–1	0	1/4	0
0	1/4	0	1/4
1	0	1/4	0

14. Lançam-se dois dados perfeitos. X indica o número obtido no primeiro dado e Y o maior ou o número comum nos dois dados.
(*a*) Escreva por meio de uma tabela de dupla entrada a distribuição conjunta de X e Y.
(*b*) As duas variáveis são independentes? Por quê?
(*c*) Calcule as esperanças e variâncias de X e Y.
(*d*) Calcule a covariância entre X e Y.
(*e*) Calcule $E(X+Y)$.
(*f*) Calcule $\text{Var}(X+Y)$.

15. Uma moeda perfeita é lançada três vezes. Sejam:
X: número de caras nos dois primeiros lançamentos;
Y: número de caras no terceiro lançamento; e
S: número total de caras.
(*a*) Usando a distribuição conjunta de (X,Y), verifique se X e Y são independentes. Qual é a covariância entre elas?
(*b*) Calcule a média e a variância das três variáveis definidas.
(*c*) Existe alguma relação entre os parâmetros encontrados em (*b*)? Por quê?

16. Depois de um tratamento, seis operários submeteram-se a um teste e, mais tarde, mediu-se a produtividade de cada um deles. A partir dos resultados apresentados na tabela ao lado, calcule o coeficiente de correlação entre a nota do teste e a produtividade.

Operário	Teste	Produtividade
1	9	22
2	17	34
3	20	29
4	19	33
5	20	42
6	23	32

17. O exemplo a seguir ilustra que $\rho = 0$ não implica independência. Suponha que (X, Y) tenha distribuição conjunta dada pela tabela abaixo.
(*a*) Mostre que $E(XY) = E(X)\,E(Y)$, donde $\rho = 0$.
(*b*) Justifique por que X e Y não são independentes.

Y \ X	−1	0	1
−1	1/8	1/8	1/8
0	1/8	0	1/8
1	1/8	1/8	1/8

8.5 Variáveis Contínuas

Nesta seção vamos considerar o caso de duas v.a. contínuas, X e Y. Nesse caso, a distribuição conjunta das duas variáveis é caracterizada por uma função $f(x, y)$, chamada *função de densidade conjunta* de X e Y, satisfazendo:

(a) $f(x, y) \geqslant 0$, para todo par (x, y);

(b) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1$;

(c) $P(a \leqslant X \leqslant b, c \leqslant Y \leqslant d) = \int_a^b\int_c^d f(x, y)dydx$.

A relação *(b)* nos diz que o volume sob a superfície representada por $f(x, y)$ é igual a 1. A relação *(c)* dá a probabilidade do par (x, y) estar num retângulo de lados b-a e d-c.

Exemplo 8.14. Suponha que $f(x, y) = 4xy$, $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$. Então, (a) está satisfeita e

$$\int_0^1\int_0^1 4xydxdy = 4\int_0^1 xdx\int_0^1 ydy = 4[x^2/2]_0^1[y^2/2]_0^1 = 1,$$

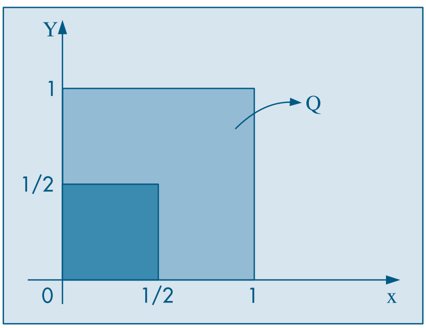
o que mostra que (b) também está satisfeita.

Calculemos $P(X \leqslant 1/2, Y \leqslant 1/2)$. A Figura 8.6 mostra o domínio de variação de X e Y e a região para a qual $X \leqslant 1/2$, $Y \leqslant 1/2$. Logo, por (c),

$$P(X \leqslant 1/2, Y \leqslant 1/2) = P(0 \leqslant X \leqslant 1/2, 0 \leqslant Y \leqslant 1/2)$$

$$= \int_0^{1/2}\int_0^{1/2} 4xydxdy = 4[x^2/2]_0^{1/2}[y^2/2]_0^{1/2} = 1/16.$$

Figura 8.6: Domínio de variação de (X, Y) para o Exemplo 8.14.



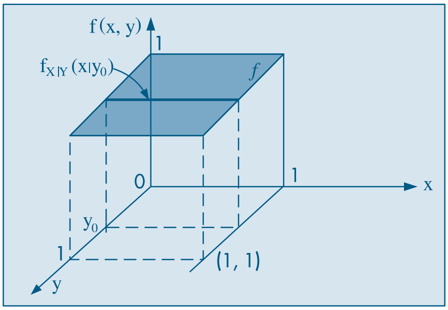
Exemplo 8.15. Suponha que a v.a. (X, Y) seja uniformemente distribuída no quadrado Q da Figura 8.6. Isso significa que

$$f(x, y) = \begin{cases} c, & \text{se } (x, y) \in Q \\ 0, & \text{caso contrário.} \end{cases} \tag{8.21}$$

Como vimos, (b) acima vale, logo $\int_0^1 \int_0^1 c dx dy = 1$ e segue-se que $c = 1$. Como a área de Q é 1, na realidade $c = \dfrac{1}{\text{área}(Q)}$. Veja a Figura 8.7.

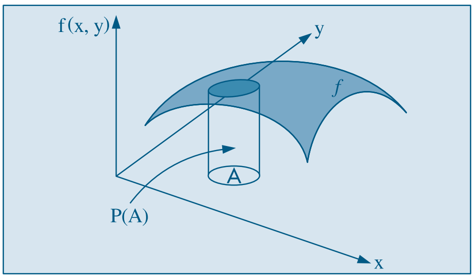
De modo geral, podemos representar a densidade bidimensional $f(x, y)$ por uma superfície no espaço tridimensional, como ilustra a Figura 8.8.

Figura 8.7: Densidade uniforme no quadrado de lado unitário, com densidade condicional representada.



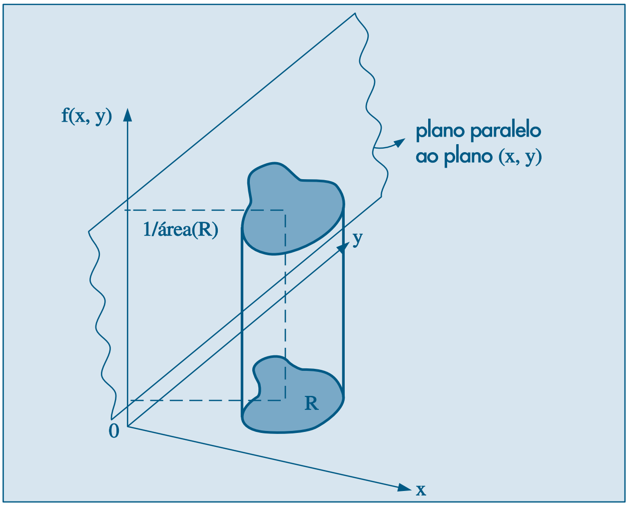
Se A for um evento, então a probabilidade $P((X, Y) \in A)$ será representada pelo volume sob a superfície, delimitado pela região A, no plano (x, y), e pela superfície cilíndrica na Figura 8.8.

Figura 8.8: Densidade como uma superfície no espaço e $P((X, Y) \in A) = P(A)$.



Se a densidade $f(x, y)$ for positiva numa região qualquer R do plano (x, y), uma v.a. diz-se *uniformemente distribuída sobre* R se $f(x, y) = 1/\text{área}(R)$, para $(x, y) \in R$, e $f(x, y) = 0$ nos demais pontos. Veja a Figura 8.9.

Figura 8.9: Distribuição uniforme na região R do plano (x, y).



Vimos que, no caso discreto, a partir da distribuição conjunta de duas v.a. X e Y, podíamos determinar a distribuição marginal de cada variável. O mesmo ocorre para v.a. contínuas.

Freqüentemente, usaremos a notação (X, Y) para denotar o par de v.a. e diremos que essa é uma v.a. bidimensional. Usamos, também, a nomenclatura *vetor bidimensional*.

Definição. Dada a v.a. bidimensional (X, Y), com função densidade de probabilidade conjunta $f(x, y)$, definimos as densidades marginais de X e Y respectivamente por

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy \tag{8.22}$$

e

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx. \tag{8.23}$$

Exemplo 8.16. Para as v.a. do Exemplo 8.14, temos

$$f_X(x) = \int_0^1 4xydy = 4x[y^2/2]_0^1 = 2x, \quad 0 \leqslant x \leqslant 1,$$

$$f_Y(y) = \int_0^1 4xydx = 2y, \quad 0 \leqslant y \leqslant 1.$$

Exemplo 8.17. Considere a v.a. (X, Y) com densidade conjunta

$$f(x, y) = \frac{2x}{y}, \quad 0 < x < 1, \quad 1 < y < e.$$

Então, as densidades marginais são dadas por

$$f_X(x) = \int_1^e \frac{2x}{y}\, dy = 2x[\ell n(y)]_1^e = 2x, \quad 0 < x < 1,$$

$$f_Y(y) = \int_0^1 \frac{2x}{y}\,dx = \frac{2}{y}\left[\frac{x^2}{2}\right]_0^1 = \frac{1}{y}, \quad 1 < y < e.$$

Para o Exemplo 8.14, vemos que o produto das densidades marginais é igual à densidade conjunta, para todo par (x, y) do domínio $[0,1] \times [0,1]$, que é o produto cartesiano dos domínios de variação de X e Y. Dizemos que as v.a. são independentes.

Definição. As variáveis aleatórias X e Y, com densidade conjunta $f(x, y)$ e marginais $f_X(x)$ e $f_y(y)$, respectivamente, são independentes se

$$f(x, y) = f_X(x)\, f_y(y), \text{ para todo par } (x, y). \tag{8.24}$$

Exemplo 8.18. Se a função densidade conjunta de X e Y for dada por

$$f(x, y) = e^{-x-y}, \quad x > 0, y > 0,$$

então é fácil ver que

$$f_X(x) = e^{-x}, \quad x > 0,$$

$$f_Y(y) = e^{-y}, \quad y > 0,$$

de modo que X e Y são independentes.

As definições de covariância, coeficiente de correlação etc. continuam, é claro, a valer para v.a. bidimensionais contínuas. Portanto, se X e Y são independentes, o coeficiente de correlação entre elas é zero.

Exemplo 8.19. Calculemos o coeficiente de correlação entre X e Y, se a densidade conjunta delas for

$$f(x, y) = x + y, \quad 0 < x < 1, \quad 0 < y < 1.$$

Temos que as marginais são dadas por

$$f_X(x) = \int_0^1 (x + y)dy = x + 1/2, \quad 0 < x < 1,$$

$$f_Y(y) = \int_0^1 (x + y)dx = y + 1/2, \quad 0 < y < 1.$$

A partir delas, calculamos médias e variâncias:

$$E(X) = \int_0^1 x(x + 1/2)dx = 7/12 = E(Y),$$

$$E(X^2) = \int_0^1 x^2(x + 1/2)dx = 5/12 = E(Y^2),$$

$$\text{Var}(X) = \text{Var}(Y) = 5/12 - 49/144 = 11/144.$$

Para calcular a covariância entre X e Y necessitamos calcular

$$E(XY) = \int_0^1\int_0^1 xy(x + y)dx\,dy = \int_0^1 (y/3 + y^2/2)\,dy = 1/3.$$

Logo,

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 1/3 - (7/12)(7/12) = -1/144.$$

Finalmente, o coeficiente de correlação entre X e Y é dado por

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)} = -\frac{1}{11}.$$

Problemas

18. As v.a. X e Y têm distribuição conjunta dada por

$$f(x,y) = \begin{cases} \frac{1}{8}x(x-y), & 0 < x < 2, -x < y < x \\ 0, & \text{caso contrário.} \end{cases}$$

(a) Faça um gráfico do domínio de variação de x e y.

(b) Prove que $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1$.

(c) Encontre as f.d.p. marginais de X e Y.

(d) Encontre a $P(X \leqslant 1)$.

19. Suponha que as v.a. X e Y tenham f.d.p.

$$f(x,y) = \begin{cases} e^{-(x+y)}, & x > 0, y > 0 \\ 0, & \text{nos demais pontos.} \end{cases}$$

(a) Calcule as f.d.p. marginais de X e Y.

(b) Calcule $P(0 < X < 1, 1 < Y < 2)$.

(c) Calcule $\rho(X, Y)$.

8.6 Distribuições Condicionais Contínuas

Nesta seção vamos tratar de obter a distribuição condicional de uma variável, dado que a outra assume um particular valor. Como sabemos, para uma v.a. contínua X, a $P(X = x) = 0$, logo a definição a seguir tem de ser interpretada apropriadamente.

Definição. A densidade condicional de X, dado que $Y = y$ é definida por

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}, \quad f_Y(y) > 0, \tag{8.25}$$

e a densidade condicional de Y, dado que $X = x$ é definida por

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}, \quad f_X(x) > 0. \tag{8.26}$$

A interpretação de (8.25), por exemplo, é a seguinte. Se $Y = y_0$, considere o plano passando por y_0 e paralelo ao plano (x, z). Esse plano determina na superfície $f(x, y) = z$ a densidade condicional $f_{X|Y}(x|y_0)$. Mesma interpretação vale para (8.26). Suponha, por exemplo, que X denote o salário de um conjunto de indivíduos e Y denote o consumo deles. Então, fixado o consumo y_0, a densidade condicional $f_{X|Y}(x|y_0)$ representa a densidade dos salários para aquele nível fixado de consumo. Nas Figuras 8.7 e 8.10 ilustramos como essa densidade condicional pode ser representada.

Exemplo 8.20. Suponha que a densidade de (X, Y) seja dada por

$$f(x, y) = 6(1 - x - y), \quad 0 < x < 1,\ 0 < y < 1 - x.$$

O domínio de variação dos pares (x, y) é o triângulo da Figura 8.11.

Figura 8.10: Densidade condicional de X, dado que $Y = y_0$.

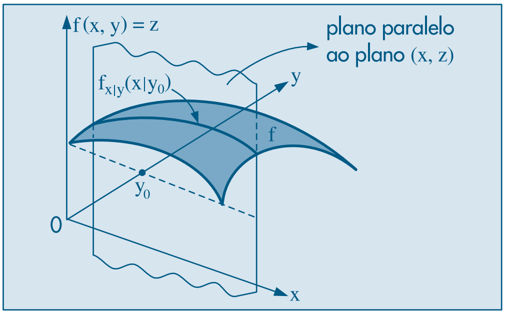
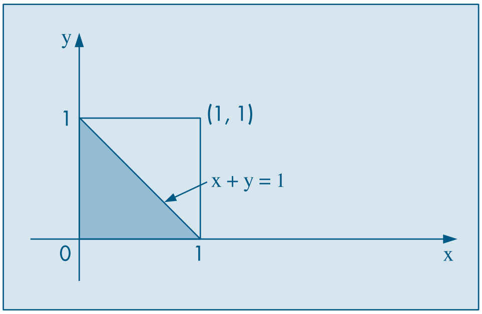


Figura 8.11: Domínio de variação de (X, Y) para o Exemplo 8.20.



Temos, então, que as densidades marginais são dadas por:

$$f_X(x) = \int_0^{1-x} 6(1 - x - y)dy = 6[y - xy - y^2/2]_0^{1-x} = 3(x - 1)^2, \quad 0 < x < 1,$$

$$f_Y(y) = \int_0^{1-y} 6(1 - x - y)dx = 3(y - 1)^2, \quad 0 < y < 1.$$

Conseqüentemente, as densidades condicionais são

$$f_{X|Y}(x|y) = \frac{2(1 - x - y)}{(y - 1)^2}, \quad 0 < x < 1 - y,$$

$$f_{Y|X}(y|x) = \frac{2(1 - x - y)}{(x - 1)^2}, \quad 0 < y < 1 - x.$$

Observe que $f_{X|Y}(x|y)$ define, de fato, uma densidade de probabilidade, para y fixado. Temos que

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)dx = \int_{-\infty}^{\infty} f(x, y)/f_Y(y)\, dx = 1/f_Y(y) \int_{-\infty}^{\infty} f(x, y)dx = f_Y(y)/f_Y(y) = 1.$$

Por exemplo, se $X = 0,5$, $f_{Y|X}(y|X = 0,5) = 4(1 - 2y)$, $0 < y < 1/2$. Essa é uma densidade que depende do valor observado de X. Assim,

$$P(0 < Y < 1/2\ |X = 0,5) = \int_0^{1/2} f_{Y|X}(y|0,5)dy = 4\int_0^{1/2} (1 - 2y)dy = 1.$$

Dado que $f_{X|Y}(x|y)$ e $f_{Y|X}(y|x)$ definem densidades de probabilidades, tem sentido em calcular suas médias, variâncias etc.

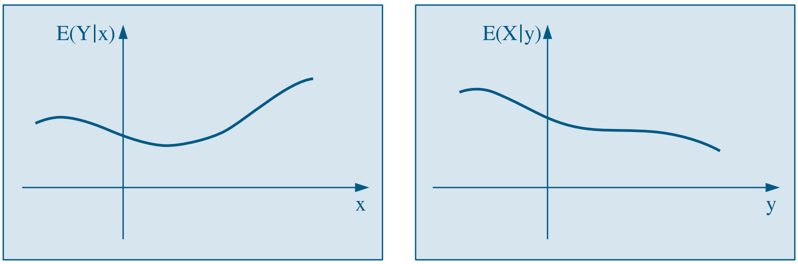
Definição. A esperança condicional de Y, dado que $X = x$, é definida por

$$E(Y|x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy, \tag{8.27}$$

e definição análoga para $E(X|y)$.

Note que $E(Y|x)$ é uma *função de* x, isto é, $E(Y|x) = s(x)$, e é denominada *curva de regressão de* Y *sobre* x. Na realidade, $E(Y|x)$ é o valor da variável aleatória $E(Y|X)$. A mesma interpretação deve ser dada para $E(X|y)$. A Figura 8.12 ilustra esses conceitos.

Figura 8.12: Curvas de regressão de Y sobre x e de X sobre y.

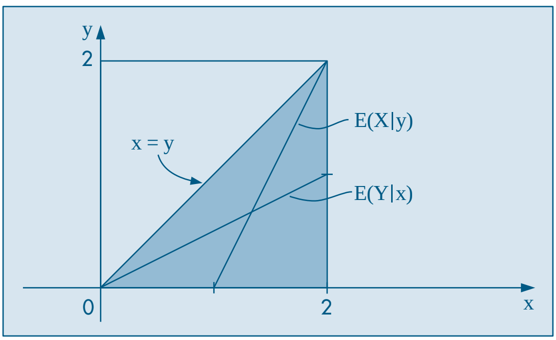


Exemplo 8.21. Suponha que

$$f(x, y) = \begin{cases} 1/2, & \text{se } x - y \geqslant 0,\ x \leqslant 2,\ x, y \geqslant 0 \\ 0, & \text{caso contrário.} \end{cases}$$

O domínio de variação de (x, y) está na Figura 8.13, juntamente com as curvas de regressão.

Figura 8.13: Curvas de regressão para o Exemplo 8.21.



Temos, então,

$$f_X(x) = \int_0^x 1/2\ dy = x/2, \quad 0 < x < 2,$$

$$f_Y(y) = \int_y^2 1/2\ dx = 1 - y/2, \quad 0 < y < 2,$$

e, portanto, as densidades condicionais são

$$f_{Y|X}(y|x) = \frac{1/2}{x/2} = 1/x, \quad 0 < y < x$$

e

$$f_{X|Y}(x|y) = \frac{1/2}{1 - y/2} = \frac{1}{2 - y}, \quad y < x < 2.$$

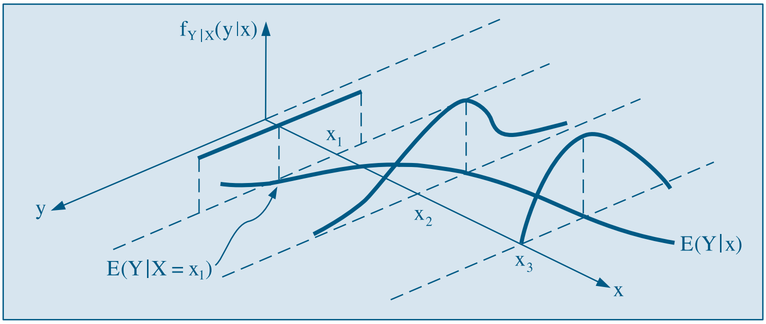
As esperanças condicionais serão dadas por

$$E(Y|x) = \int_0^x y\ \frac{1}{x}\ dy = \frac{x}{2},$$

$$E(X|y) = \int_y^2 x\ \frac{1}{2 - y}\ dx = 1 + \frac{y}{2}.$$

Note, portanto, que ambas as curvas de regressão são *funções lineares*, como ilustra a Figura 8.13. No caso geral, a Figura 8.14 mostra como seriam essas médias condicionais.

Figura 8.14: Representação gráfica da curva de regressão de Y sobre x.



Observe, também, que se, por exemplo, $X = 1$, $E(Y|1) = 1/2$.

Problemas

20. Calcule $f_{X|Y}(x|y)$ e $f_{Y|X}(y|x)$ para a densidade do Problema 18.

21. Calcule as densidades condicionais para o Problema 19. Comente.

22. Calcule as densidades marginais e condicionais para a v.a. (X, Y), com f.d.p.

$$f(x, y) = (1/64)(x + y), \quad 0 \leqslant x \leqslant 4, \ 0 \leqslant y \leqslant 4.$$

23. Mesmos itens do Problema 22 para a f.d.p. conjunta

$$f(x, y) = 3e^{-(x+3y)}, \quad x > 0, y > 0.$$

24. Calcule as esperanças condicionais $E(Y|x)$ e $E(X|y)$ para o Problema 21.

25. Calcule as esperanças condicionais para o Problema 22.

26. Prove que $E(E(X|Y)) = E(X)$.

(Sugestão: $E(X|y)$ é uma função de y e portanto é uma v.a. Na realidade, $E(X|y)$ é o valor da v.a. $E(X|Y)$! Considere a expressão para $E(X|y)$ e tome a esperança novamente. Mude a ordem das integrais e obtenha o resultado.)

8.7 Funções de Variáveis Contínuas

O tratamento desta seção é uma extensão daquele para uma variável contínua (ver seção 7.6). Considere duas variáveis X e Y, com função densidade conjunta $f(x, y)$ e suponha que queremos obter a densidade das variáveis Z e W, tais que

$$Z = h_1(X, Y)$$
$$W = h_2(X, Y)$$

Suponha que possamos expressar x e y em função de z e w, isto é,

$$x = g_1(z, w),$$
$$y = g_2(z, w).$$

Supondo que as derivadas parciais de x e y, em relação a z e w, existam e sejam contínuas, podemos obter a densidade conjunta de Z e W através de

$$g(z, w) = f(g_1(z, w), g_2(z, w))|J|, \tag{8.28}$$

onde J é o Jacobiano da transformação que leva (x, y) em (z, w), dado por

$$J = \begin{vmatrix} \dfrac{\partial x}{\partial z} & \dfrac{\partial x}{\partial w} \\ \dfrac{\partial y}{\partial z} & \dfrac{\partial y}{\partial w} \end{vmatrix}$$

No caso unidimensional, $Y = h(X)$, J era simplesmente $\dfrac{dx}{dy}$, com $x = h^{-1}(y)$.

Exemplo 8.22. Retomemos o Exemplo 8.14, no qual tínhamos

$$f_X(x) = 2x, \quad 0 < x < 1,$$

$$f_Y(y) = 2y, \quad 0 < y < 1,$$

e X e Y eram independentes.

Suponha que queiramos determinar a densidade $F_Z(z)$ da v.a. $Z = XY$. Considere $W = X$ e portanto $x = w$, $y = \dfrac{z}{w}$ e o Jacobiano é

$$\begin{vmatrix} 0 & 1 \\ \dfrac{1}{w} & \dfrac{-z}{w^2} \end{vmatrix} = -\frac{1}{w},$$

de modo que

$$g(z, w) = 4w \frac{z}{w} \left| -\frac{1}{w} \right| = \frac{4z}{w}, \quad 0 < w < 1, 0 < \frac{z}{w} < 1.$$

Segue-se que $0 < z < w < 1$ e a densidade de Z é obtida por

$$f_Z(z) = \int_z^1 g(z, w)dw = \int_z^1 \frac{4z}{w} dw = -4z\, \ell n(z), \quad 0 < z < 1.$$

Problemas

27. Encontre a densidade de $Z = X + Y$ para X e Y v.a. independentes, com $f_X(x) = 2x$, $0 < x < 1$ e $f_Y(y) = 2y$, $0 < y < 1$.
(Sugestão: considere $0 < z < 1$ e $1 < z < 2$.)

28. Se X tiver densidade $f_X(x) = 2x$, $0 \leqslant x \leqslant 1$ e Y tiver densidade $f_Y(y) = y^2/9$, $0 \leqslant y \leqslant 3$ e forem independentes, encontre a densidade de $W = XY$.

29. Encontre a densidade de $Z = X/Y$, se X e Y são independentes, com densidades $f_X(x) = e^{-x}$, $x > 0$ e $f_Y(y) = 2e^{-2y}$, $y > 0$.
(Sugestão: $z = x/y$, $w = y$.)

8.8 Distribuição Normal Bidimensional

Assim como a distribuição normal é um modelo importante para variáveis contínuas unidimensionais, para v.a. contínuas bidimensionais podemos considerar o modelo normal bidimensional, definido a seguir.

Definição. A variável (X, Y) tem distribuição normal bidimensional se sua densidade conjunta for dada por

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}, \tag{8.29}$$

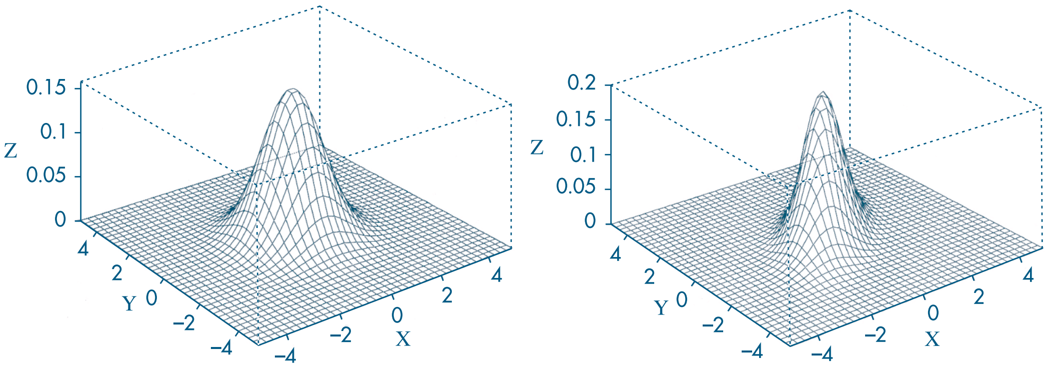
para $-\infty < x < \infty$, $-\infty < y < \infty$.

Aqui, estamos usando a notação $\exp\{z\} = e^z$.

Vemos que a densidade em questão depende de cinco parâmetros: as médias μ_x e μ_y, que podem assumir quaisquer valores reais, as variâncias σ_x^2 e σ_y^2, que devem ser positivas, e o coeficiente de correlação ρ entre X e Y, que deve satisfazer $-1 < \rho < 1$.

Dois exemplos de gráficos dessa densidade estão representados na Figura 8.15.

Figura 8.15: f.d.p. de normais bidimensionais.
(a) $\mu_x = \mu_y = 0$, $\sigma_x = \sigma_y = 1$, $\rho = 0$; (b) $\mu_x = \mu_y = 0$, $\sigma_x = \sigma_y = 1$, $\rho = 0{,}6$.



As seguintes propriedades podem ser demonstradas:

(a) As distribuições marginais de X e Y são normais unidimensionais, a saber

$$X \sim N(\mu_x, \sigma_x^2), \quad Y \sim N(\mu_y, \sigma_y^2).$$

(b) $\rho = \text{Corr}(X, Y)$.

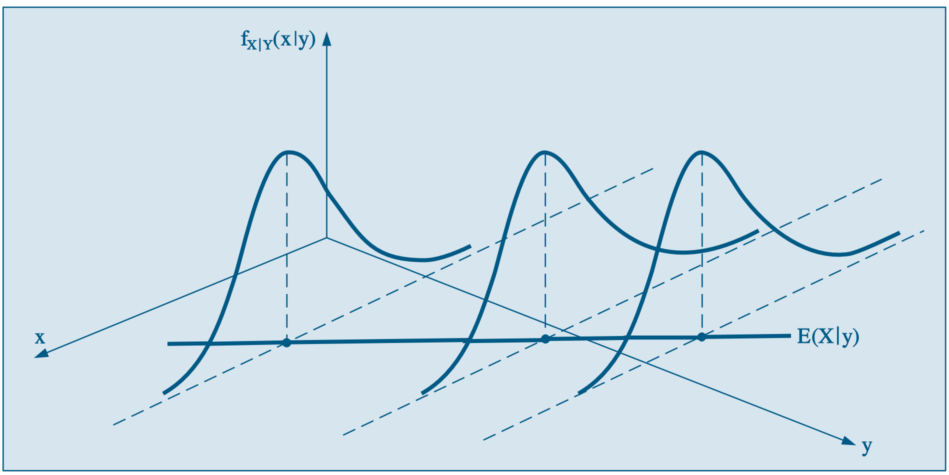
(c) As distribuições condicionais são normais, com

$$f_{Y|X}(y|x) \sim N\left(\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x),\ \sigma_y^2(1 - \rho^2)\right),$$

$$f_{X|Y}(x|y) \sim N\left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y),\ \sigma_x^2(1 - \rho^2)\right).$$

Ou seja, as médias condicionais são funções lineares. Ver Figura 8.16.

Figura 8.16: Curva de regressão de X sobre y para o caso da normal bidimensional.



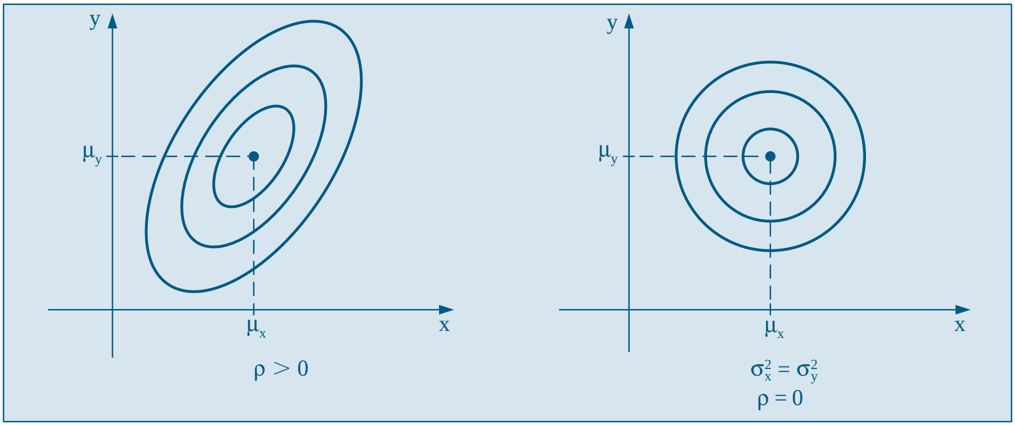
Se chamarmos $z = f(x, y)$, então $z = c$, constante, determina sobre a superfície uma *curva* de *nível*, que nesse caso é uma elipse. Variando c, teremos as diversas curvas de nível (que são curvas onde a densidade de probabilidade é constante), semelhantes às curvas de nível de um mapa de relevo. No caso em que $\rho = 0$ e as variâncias são iguais, isto é, $\sigma^2_x = \sigma^2_y$, essas curvas serão círculos. Veja a Figura 8.17

Vimos que $\rho = 0$ significa que as variáveis X e Y são não-correlacionadas. Aqui, poderemos concluir algo mais. Nessa situação poderemos escrever a densidade (8.29) como

$$f(x, y) = \left(\frac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}\right)\left(\frac{1}{\sigma_y\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}\right), \tag{8.30}$$

isto é, a densidade conjunta é o produto das duas marginais, que sabemos serem normais. Ou seja, concluímos que X e Y são independentes. Portanto, no caso em que X e Y tiverem densidade conjunta normal bivariada, $\rho = 0$ é equivalente à independência entre X e Y.

Figura 8.17: Curvas de nível para a normal bidimensional.



8.9 Problemas e Complementos

30. Um *sinal* consiste numa série de vibrações de magnitude X, tendo os valores -1, 0, 1, cada um com probabilidade 1/3. Um *ruído* consiste numa série de vibrações, de magnitude Y, tendo os valores -2, 0, 2, com probabilidades 1/6, 2/3, 1/6, respectivamente. Combinando-se o sinal com o ruído, obtemos o sinal efetivamente observado, $Z = X + Y$. Construa a função de probabilidade para Z e calcule a sua média e variância, admitindo que sinal e ruído são independentes.

31. Numa comunidade em que apenas dez casais trabalham, fez-se um levantamento no qual foram obtidos os seguintes valores para os rendimentos anuais: