Capítulo 4

Variáveis aleatórias

Neste capítulo, introduzimos as variáveis aleatórias e suas distribuições de probabilidade.

Definição 4.1 *Dado um experimento aleatório, descrito pelo espaço de probabilidades* $(\Omega, \mathcal{E}, \mathbb{P})$*, uma função numérica* $X : \Omega \to \mathbb{R}$ *será dita uma variável aleatória (do experimento).*

Exemplo 4.2 *No Exemplo 3.1,* $X =$ *"número lançado" é uma variável aleatória. Mais precisamente,* $X : \Omega = \{1, 2, \ldots, 6\} \to \mathbb{R}$ *tal que* $X(\omega) = \omega$ *é uma função numérica do experimento, e logo é uma variável aleatória.*

Exemplo 4.3 *No Exemplo 3.2,* $X =$ *"número escolhido" é uma variável aleatória. Neste caso,* $X : \Omega = [0, 1] \to \mathbb{R}$ *tal que* $X(\omega) = \omega$.

Exemplo 4.4 *No Exemplo 3.5,* $X =$ *"número de lançamentos" é uma variável aleatória. Neste caso,* $X : \Omega \to \mathbb{R}$ *tal que* $X(\omega) = \omega$.

Exemplo 4.5 *No Exemplo 3.7,* $X =$ *"soma dos números lançados" é uma variável aleatória. Neste caso,* $X : \Omega = \{(i, j) : i, j = 1, 2, \ldots, 6\} \to \mathbb{R}$ *tal que* $X((i, j)) = i + j$.

Exemplo 4.6 *No Exemplo 3.6,* $X =$ *"distância do ponto escolhido à origem" é uma variável aleatória. Neste caso,* $X : \Omega = \mathcal{C} \to \mathbb{R}$ *tal que* $X((x, y)) = \sqrt{x^2 + y^2}$.

Exemplo 4.7 (Amostragem aleatória) *De volta ao contexto do parágrafo sobre amostragem aleatória em populações (começando na página 91), seja* Π *uma população e* $X : \Pi \to \mathbb{R}$ *uma variável (numérica) definida em* Π *(neste contexto, para diferenciar de variável aleatória, falaremos em* variável populacional*). Agora, consideremos uma amostragem casual simples de tamanho 1 nesta população. Como visto no parágrafo mencionado acima, o espaço amostral é* $\Omega = \Pi$*. Podemos considerar então o valor da variável* X *no indivíduo sorteado. Neste contexto,* $X : \Omega \to \mathbb{R}$ *é uma variável aleatória. No Exemplo 3.8,* $X = 1_{mulher}$ *é a variável (aleatória) indicadora de mulher.*

Exemplo 4.8 (Amostragem casual simples com/sem reposição) *No Exemplo 3.9,* $X =$ *"número de mulheres na amostra" é uma variável aleatória. Podemos considerar o caso sem reposição também.*

4.1 Distribuição de probabilidades de variáveis aleatórias

O que nos interessa nas variáveis aleatórias são suas distribuições de probabilidade, isto é, as probabilidades dos diversos eventos envolvendo tais variáveis. Como no caso das variáveis populacionais, temos o caso discreto e o caso contínuo.

No Exemplo 4.2, os valores possíveis de X perfazem o conjunto $\{1, \dots, 6\}$. Seguem exemplos de eventos envolvendo a v.a. X.

$$\{X = 1\} = \{\omega : X(\omega) = 1\} = \{1\},\ \{X = 2\} = \{2\},$$
$$\{X \leq 2\} = \{\omega : X(\omega) \leq 2\} = \{1, 2\},\ \{X \geq 3\} = \{3, 4, 5\}$$

Então,

$$\mathbb{P}(\{X = 1\}) =^{1} \mathbb{P}(X = 1) = \mathbb{P}(\{1\}) = \tfrac{1}{6} = \mathbb{P}(X = 2)$$
$$\mathbb{P}(X \leq 2) = \mathbb{P}(\{1, 2\}) = \tfrac{1}{3},\ \mathbb{P}(X \geq 3) = \mathbb{P}(\{3, 4, 5\}) = \tfrac{1}{2}.$$

[1]Omitiremos daqui para frente as chaves dos eventos envolvendo variáveis aleatórias dentro do sinal de probabilidade.

x	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Tabela 4.1

4.1.1 Variáveis aleatórias discretas

Quando o conjunto de valores possíveis de uma v.a. X for finito ou infinito enumerável, como no exemplo acima, em que ele é finito, dizemos que X é discreta. Neste caso, sendo $\mathcal{V}_X = \{x_i,\ i = 1, 2, \ldots\}$ o conjunto de valores, então se tivermos as probabilidades de todos os eventos $\{X = x_i\}$, $i = 1, 2, \ldots$, (que diremos unitários), então, pela aditividade da probabilidade (3.20), podemos obter as probabilidades de eventos compostos como $\{X \leq w\}$, $\{X > z\}$, onde w, z são números arbitrários, e outros, como segue.

$$\mathbb{P}(X \leq w) = \sum_{i: x_i \leq w} \mathbb{P}(X = x_i), \quad \mathbb{P}(X > z) = \sum_{i: x_i > z} \mathbb{P}(X = x_i)$$

A distribuição de probabilidades de X é pois determinada pelas probabilidades dos eventos unitários, ou em outras palavras pela *função de probabilidade* $\mathbb{P}(X = \cdot) : \mathcal{V}_X \to [0, 1]$.

No Exemplo 4.2, a distribuição (de probabilidades) de X é pois dada por

$$\mathbb{P}(X = i) = \frac{1}{6}, \quad i = 1, \ldots, 6. \tag{4.1}$$

No Exemplo 4.4, temos $\mathcal{V}_X = \{1, 2, \ldots\}$, os inteiros positivos, e, como já vimos no Exemplo 3.5, $\mathbb{P}(X = i) = \mathbb{P}(\{i\}) = 2^{-i}$, $i \in \mathcal{V}_X$.

No Exemplo 4.5, $\mathcal{V}_X = \{2, 3, \ldots, 12\}$, e

$$\begin{aligned}
\{X = 2\} &= \{(1, 1)\} \\
\{X = 3\} &= \{(1, 2), (2, 1)\} \\
\{X = 4\} &= \{(1, 3), (2, 2), (3, 1)\} \\
\{X = 5\} &= \{(1, 4), (2, 3), (3, 2), (4, 1)\}
\end{aligned}$$

e assim por diante, de forma que, lembrando que se trata de um espaço equiprovável, podemos representar a função de probabilidade de X na Tabela 4.1.

4.1.2 Variáveis aleatórias contínuas

No caso em que $\mathcal{V}_X$ for um conjunto infinito não enumerável, em geral não é suficiente obter as probabilidades dos eventos unitários $\{X = x\}$, $x \in \mathcal{V}_X$ (nos casos que veremos nestas notas, estas probabilidades se anulam todas). Vejam a discussão feita no parágrafo sobre espaços amostrais não enumeráveis (página 86).

Neste caso, para caracterizar a distribuição de probabilidades de X é suficiente termos as probabilidades dos eventos $\{X \in I\}$, onde I é um intervalo arbitrário da reta. Nos casos que veremos nestas notas, tais probabilidades serão dadas por *funções de densidade de probabilidade* f_X. Isto é, existirá uma função $f_X : \mathbb{R} \to [0, \infty)$ tal que

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)\, dx$$

para todo $a, b \in \mathbb{R}$ com $a \leq b$.

No Exemplo 4.3, $\mathbb{P}(a < X < b) = (b \wedge 1) - (a \vee 0)$, onde $\vee$ indica o máximo e $\wedge$ indica o mínimo. Concluímos que para $f = 1_{[0,1]}$, a função indicadora do intervalo $[0, 1]$, isto é

$$1_{[0,1]}(x) = \begin{cases} 1, & \text{se } 0 \leq x \leq 1 \\ 0,, & \text{caso contrário,} \end{cases}$$

temos $\mathbb{P}(a < X < b) = \int_a^b f(x)\, dx$, e então f é a função de densidade de probabilidade de X.

Na descrição da distribuição de uma v.a. contínua, é suficiente considerarmos intervalos I semiinfinitos $(-\infty, a]$.

No Exemplo 4.6, temos que

$$\mathbb{P}(X \leq a) = \begin{cases} 1, & \text{se } a \geq 1, \\ \frac{\pi a^2}{\pi} = a^2, & \text{se } 0 \leq a \leq 1, \\ 0, & \text{se } a \leq 0, \end{cases}$$

logo $f(x) = 2x\, 1_{[0,1]}(x)$ é a função de densidade de probabilidade de X neste caso (verifique).

Observação 4.9 *Como objetos matemáticos, as funções de probabilidade e funções de freqüência, de um lado, e as funções de densidade de probabilidade e funções de densidade de freqüência, por outro, são idênticas, respectivamente, isto é, são todas funções não negativas satisfazendo (1.7) e (1.14), respectivamente. Uma situação em que estes objetos se identificam é a seguinte.*

Observação 4.10 *Seja* $X : \Pi \to \mathbb{R}$ *uma variável populacional definida na população* Π*, e façamos uma amostragem casual simples de tamanho* 1 *em* Π*. Como vimos no Exemplo 4.7,* X *observada no indivíduo amostrado é uma variável aleatória. Qual é a distribuição de probabilidades de* X*?*

Vamos supor que X *seja uma variável populacional discreta, cuja distribuição de freqüências é dada pela função de freqüência* $P(X = \cdot)$*. Então para* $x \in \mathcal{V}_X$*,*

$$P(X = x) = \frac{\#\{I \in \Pi : X(I) = x\}}{\#\Pi}. \tag{4.2}$$

Por outro lado, a probabilidade do evento $\{X = x\}$ *é dada por*

$$\mathbb{P}(X = x) = \frac{\#\{X = x\}}{\#\Omega}, \tag{4.3}$$

pois se trata de espaço equiprovável. Mas como re/vimos no Exemplo 4.7, $\Omega = \Pi$*, e logo os lados direitos de (4.2) e (4.3) são iguais. Portanto,*

$$P(X = x) = \mathbb{P}(X = x)$$

para $x \in \mathcal{V}_X$*, e temos a coincidência das funções de freqüência e de probabilidade de* X*, vista como variável populacional e aleatória, respectivamente.*

Por um raciocínio análogo, valendo-nos neste caso de aproximações, concluimos que também no caso de uma variável populacional contínua X*, se fizermos amostragem casual simples de tamanho* 1 *na população em questão, e observarmos* X *no indivíduo sorteado, então a distribuição de probabilidades de* X*, variável aleatória neste contexto, é dada por uma função de densidade de probabilidade, que é idêntica à função de densidade de freqüência de* X *vista como variável populacional.*

Em conclusão, ao fazermos uma amostragem casual simples de tamanho 1 de variável populacional, obtemos uma variável aleatória, cuja distribuição de probabilidades é dada pela distribuição de freqüências da variável populacional. Isto se manifesta, em particular, na coincidência comentada na Observação 4.9 acima.

$x \in$	$(-\infty,1)$	$[1,2)$	$[2,3)$	$[3,4)$	$[4,5)$	$[5,6)$	$[6,+\infty)$
$F_X(x)$	0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	1

Tabela 4.2

Observação 4.11 *Em vista da Observação 4.9, é natural empregarmos as mesmas formas de descrição para as distribuições de probabilidades de variáveis aleatórias do que as utilizadas para as distribuições de freqüências de variáveis populacionais. Fazemos isto nas subseções seguintes.*

4.1.3 Função de distribuição acumulada

Dada uma variável aleatória X, sua *função de distribuição (acumulada)* F_X é a função $F_X : \mathbb{R} \to [0,1]$ tal que para todo $x \in \mathbb{R}$

$$F_X(x) = \mathbb{P}(X \leq x).$$

(Compare com a definição no início da Subseção 1.3.)

Como no caso das funções de distribuição de variáveis populacionais, a função de distribuição é não decrescente, satisfazendo

$$\lim_{x\to-\infty} F_X(x) = 0, \qquad \lim_{x\to+\infty} F_X(x) = 1.$$

No caso de v.a.'s discretas, a função de distribuição é do tipo escada (constante por partes, com saltos; veja a Observação 1.17). Para v.a.'s contínuas, a função de distribuição é contínua (veja a Observação 1.18).

No Exemplo 4.2, F_X é representada na Tabela 4.2, e seu gráfico aparece na Figura 4.1.

No caso do Exemplo 4.3, vimos acima que $f = 1_{[0,1]}$ é a função de densidade de probabilidade de X, logo

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0; \\ \int_0^x dy = x, & \text{if } 0 \leq x \leq 1; \\ 1, & \text{if } x > 1, \end{cases} \tag{4.4}$$

cujo gráfico aparece na Figura 4.2.

Observação 4.12 *Como no caso das distribuições de freqüências para variáveis populacionais, a função de distribuição de uma variável aleatória determina sua distribuição de probabilidades. (Veja a Observação 1.19.)*

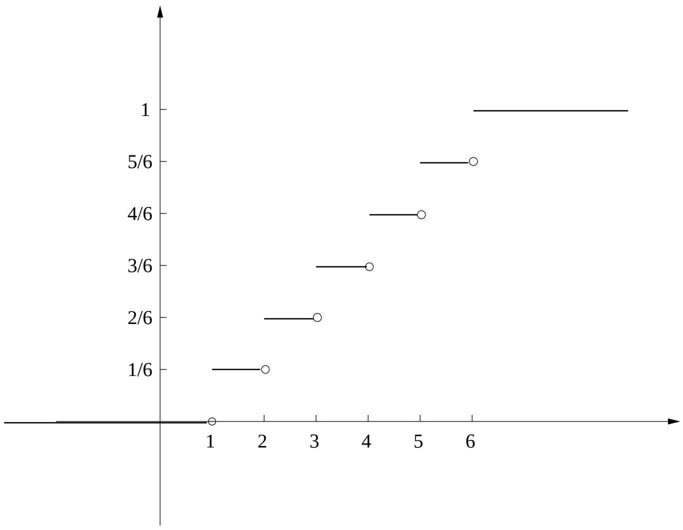

Figura 4.1

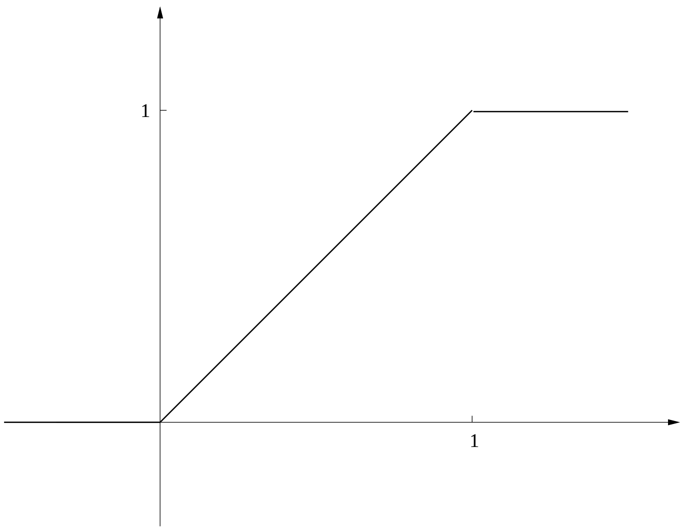

Figura 4.2

4.1.4 Esperança

A *esperança* de uma variável aleatória é por definição a média de sua distribuição de probabilidades (no mesmo sentido de média de distribuição de freqüências vista na Subseção 2.1.1). Isto é, se X for uma v.a. discreta, então

$$\mathbb{E}(X) = \sum_x x\,\mathbb{P}(X = x), \tag{4.5}$$

onde a soma é sobre o conjunto $\mathcal{V}_X$ de valores de X, e $\mathbb{P}(X = \cdot)$ é a função de probabilidade de X; e se X for uma v.a. contínua, então,

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx, \tag{4.6}$$

onde f_X é a função de densidade de probabilidade de X.

Outros nomes usados para designar a esperança são *valor esperado* e *média*.

No Exemplo 4.2, de (4.1), temos que

$$\mathbb{E}(X) = \sum_{i=1}^{6} i\,\mathbb{P}(X = i) = \frac{1}{6}\sum_{i=1}^{6} i = \frac{1}{6}\frac{6\times 7}{2} = \frac{21}{6} = 3.5. \tag{4.7}$$

No Exemplo 4.3, como já vimos $f_X = 1_{[0,1]}$, e logo temos que

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \int_0^1 x\, dx = \frac{1^2 - 0^2}{2} = \frac{1}{2}. \tag{4.8}$$

Observação 4.13 *Matematicamente, a esperança de uma variável aleatória é idêntica à média de uma variável populacional, representando o centro de massa da distribuição de probabilidade da variável aleatória. As propriedades matemáticas da esperança de uma v.a. são portanto as mesmas da média de uma variável populacional. A saber, dada uma função $h : \mathbb{R} \to \mathbb{R}$, temos*

$$\mathbb{E}[h(X)] = \sum_x h(x)\,\mathbb{P}(X = x), \tag{4.9}$$

se X for discreta, e

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(x)\, f_X(x)\, dx, \tag{4.10}$$

se X for contínua.

De onde segue a linearidade *da esperança: dadas constantes numéricas a, b, temos que*

$$\mathbb{E}(a + b\,X) = a + b\,\mathbb{E}(X). \tag{4.11}$$

(Veja a Proposição 2.7 e o Corolário 2.10. Formas mais gerais das duas propriedades acima, envolvendo mais de uma v.a., serão vistas adiante, na Seção 4.4.)

Observação 4.14 *Adiante, na Seção 4.4 (veja a Observação 4.46), veremos uma interpretação* estatística *para a esperança de uma v.a., que em particular justifica o termo.*

Terminamos a subseção com uma forma alternativa de expressar a esperança no caso de variáveis não negativas.

Proposição 4.15 *Suponha que X seja uma v.a. inteira e não negativa. Então,*

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i), \tag{4.12}$$

$$\mathbb{E}(X^2) = 2\sum_{i=1}^{\infty} i\,\mathbb{P}(X \geq i) - \mathbb{E}(X). \tag{4.13}$$

Demonstração Da definição ((4.5), já que se trata de v.a. discreta),

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{i=1}^{\infty} i\,\mathbb{P}(X = i) = \sum_{i=1}^{\infty}\sum_{j=1}^{i} \mathbb{P}(X = i) = \sum_{1\leq j\leq i<\infty} \mathbb{P}(X = i) \\
&= \sum_{j=1}^{\infty}\sum_{i=j}^{\infty} \mathbb{P}(X = i) = \sum_{j=1}^{\infty} \mathbb{P}(X \geq j), \qquad (4.14)
\end{aligned}$$

e (4.12) está provado. De (4.9)

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{i=1}^{\infty} i^2\,\mathbb{P}(X=i) = \sum_{i=1}^{\infty} i^2\,[\mathbb{P}(X\geq i) - \mathbb{P}(X\geq i+1)]\\
&= \sum_{i=1}^{\infty} i^2\,\mathbb{P}(X\geq i) - \sum_{i=1}^{\infty} i^2\,\mathbb{P}(X\geq i+1)\\
&= \sum_{i=1}^{\infty} i^2\,\mathbb{P}(X\geq i) - \sum_{i=2}^{\infty} (i-1)^2\,\mathbb{P}(X\geq i)\\
&= \mathbb{P}(X\geq 1) + \sum_{i=2}^{\infty}[i^2-(i-1)^2]\mathbb{P}(X\geq i)\\
&= \sum_{i=1}^{\infty}(2i-1)\mathbb{P}(X\geq i) = 2\sum_{i=1}^{\infty} i\,\mathbb{P}(X\geq i) - \sum_{i=1}^{\infty}\mathbb{P}(X\geq i)\\
&\stackrel{(4.12)}{=} 2\sum_{i=1}^{\infty} i\,\mathbb{P}(X\geq i) - \mathbb{E}(X), \qquad (4.15)
\end{aligned}$$

e (4.13) está provado.

Observação 4.16 *(4.12,4.13) são equivalentes a*

$$\mathbb{E}(X) = \sum_{i=0}^{\infty}\mathbb{P}(X>i), \qquad (4.16)$$

$$\mathbb{E}(X^2) = 2\sum_{i=0}^{\infty} i\,\mathbb{P}(X>i) + \mathbb{E}(X). \qquad (4.17)$$

A Proposição 4.15 tem a seguinte versão contínua, cuja demonstração se baseia na *integração por partes* do Cálculo, e por isto será omitida.

Proposição 4.17 *Suponha que X seja uma v.a. contínua e não negativa. Então,*

$$\mathbb{E}(X) = \int_0^{\infty}\mathbb{P}(X>x)\,dx, \qquad (4.18)$$

$$\mathbb{E}(X^2) = 2\int_0^{\infty} x\,\mathbb{P}(X>x)\,dx. \qquad (4.19)$$

4.1.5 Variância

A *variância* de uma variável aleatória é por definição a variância de sua distribuição de probabilidades (no mesmo sentido de variância de distribuição de freqüências vista na Subseção 2.2.1). Isto é,

$$\mathbb{V}(X) = \mathbb{E}\{[X - \mathbb{E}(X)]^2\}. \tag{4.20}$$

Se X for uma v.a. discreta, então

$$\mathbb{V}(X) = \sum_x [x - \mathbb{E}(X)]^2 \, \mathbb{P}(X = x), \tag{4.21}$$

onde a soma é sobre o conjunto $\mathcal{V}_X$ de valores de X, e $\mathbb{P}(X = \cdot)$ é a função de probabilidade de X; e se X for uma v.a. contínua, então,

$$\mathbb{V}(X) = \int_{-\infty}^{\infty} [x - \mathbb{E}(X)]^2 \, f_X(x) \, dx, \tag{4.22}$$

onde f_X é a função de densidade de probabilidade de X.

O *desvio-padrão* de X é a raiz quadrada da variância.

$$\mathbb{DP}(X) = \sqrt{\mathbb{V}(X)}. \tag{4.23}$$

Variância e desvio padrão para variáveis aleatórias têm as mesmas propriedades do que no caso de variáveis populacionais (já que se trata de objetos matemáticos idênticos). Temos a fórmula alternativa:

$$\mathbb{V}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2, \tag{4.24}$$

e, para a, b constantes numéricas,

$$\mathbb{V}(a + b\,X) = b^2\,\mathbb{V}(X), \quad \mathbb{DP}(a + b\,X) = |b|\,\mathbb{DP}(X). \tag{4.25}$$

(Veja (2.80), (2.93), (2.94).)

No Exemplo 4.2, de (4.1) e (4.9), temos que

$$\mathbb{E}(X^2) = \sum_{i=1}^{6} i^2\,\mathbb{P}(X = i) = \frac{1}{6}\sum_{i=1}^{6} i^2 = \frac{1}{6}\frac{13 \times 6 \times 7}{6} = \frac{91}{6} = 15.17. \tag{4.26}$$

De (4.24) e (4.7), temos

$$\mathbb{V}(X) = 15.17 - (3.5)^2 = 15.17 - 12.25 = 2.92; \quad \mathbb{DP}(X) = \sqrt{2.92} = 1.71. \tag{4.27}$$

No Exemplo 4.3, temos que

$$\mathbb{E}(X^2) = \int_0^1 x^2\, dx = \frac{1}{3} \tag{4.28}$$

De (4.24) e (4.8), temos

$$\mathbb{V}(X) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12} = 0.08; \quad \mathbb{DP}(X) = 0.28. \tag{4.29}$$

4.2 Modelos para variáveis aleatórias discretas

Nesta seção apresentamos alguns modelos para variáveis aleatórias discretas. O foco é na distribuição de probabilidades e suas propriedades, como esperança e variância. Em alguns casos, ela será deduzida de descrição de experimento aleatório subjacente.

4.2.1 O modelo uniforme

Seja $\mathcal{V}$ um subconjunto finito da reta. Dizemos que a v.a. X tem distribuição *uniforme* em $\mathcal{V}$ (notação: $X \sim \mathrm{U}(\mathcal{V})$) se

$$\mathbb{P}(X = x) = \frac{1}{\#\mathcal{V}},\ x \in \mathcal{V}. \tag{4.30}$$

Vamos calcular a esperança e variância de $X \sim \mathrm{U}(\{1, 2, \ldots, N\})$. Um caso particular de uma v.a. com esta distribuição é aquela do Exemplo 4.2, com $N = 6$.

$$\mathbb{E}(X) = \sum_{i=1}^{N} i\,\frac{1}{N} = \frac{1}{N}\sum_{i=1}^{N} i = \frac{1}{N}\frac{N(N+1)}{2} = \frac{N+1}{2}, \tag{4.31}$$

$$\mathbb{E}(X^2) = \frac{1}{N}\sum_{i=1}^{N} i^2 = \frac{1}{N}\frac{(2N+1)N(N+1)}{6} = \frac{(2N+1)(N+1)}{6}, \tag{4.32}$$

e logo, de (4.24),

$$
\begin{aligned}
\mathbb{V}(X) &= \frac{(2N+1)(N+1)}{6} - \left(\frac{N+1}{2}\right)^2 \\
&= \frac{N+1}{12}[2(2N+1) - 3(N+1)] \\
&= \frac{(N+1)(N-1)}{12} = \frac{N^2-1}{12}. \qquad (4.33)
\end{aligned}
$$

4.2.2 O modelo de Bernoulli

Dado $p \in [0,1]$, dizemos que a v.a. X tem distribuição de *Bernoulli* com parâmetro p (notação: $X \sim$ Ber(p)) se $\mathcal{V}_X = \{0,1\}$ e

$$\mathbb{P}(X=1) = p, \quad \mathbb{P}(X=0) = 1-p. \qquad (4.34)$$

Esta v.a. surge naturalmente em experimentos aleatórios da seguinte forma. Seja $(\Omega, \mathcal{E}, \mathbb{P})$ um espaço de probabilidades para um experimento aleatório, e $A \in \mathcal{E}$ um evento deste espaço. Seja $X = 1_A$ a v.a. *indicadora* de A, isto é,

$$X(\omega) = 1_A(\omega) = \begin{cases} 1, & \text{se } \omega \in A, \\ 0, & \text{se } \omega \in A^c. \end{cases} \qquad (4.35)$$

Então $X \sim$ Ber(p), onde $p = \mathbb{P}(A)$.

Vamos calcular a esperança e variância de $X \sim$ Ber(p).

$$
\begin{aligned}
\mathbb{E}(X) &= 0 \times \mathbb{P}(X=0) + 1 \times \mathbb{P}(X=1) = 0 \times (1-p) + 1 \times p = p \\
&= 0^2 \times \mathbb{P}(X=0) + 1^2 \times \mathbb{P}(X=1) = \mathbb{E}(X^2) \qquad (4.36)
\end{aligned}
$$

e logo, de (4.24),

$$\mathbb{V}(X) = p - p^2 = p(1-p). \qquad (4.37)$$

4.2.3 O modelo binomial

Como no modelo anterior sejam $(\Omega, \mathcal{E}, \mathbb{P})$ um espaço de probabilidades para um experimento aleatório, e $A \in \mathcal{E}$ um seu evento. Seja $p = \mathbb{P}(A)$ e considere $n \geq 1$ realizações independentes do experimento em questão, sob as mesmas condições, e seja X o número de realizações em que A ocorre.

Vamos calcular a distribuição de X. Em primeiro lugar, é claro que $\mathcal{V}_X = \{0, 1, \ldots, n\}$.

Para $k \in \mathcal{V}_X$, o evento $\{X = k\}$ consiste de todas as seqüências de n realizações do experimento em que A ocorre exatamente k vezes, e logo A^c ocorre exatamente $n - k$ vezes. Pela independência e igualdade de condições entre as realizações do experimento, cada uma das seqüências mencionadas acima têm a mesma probabilidade $p^k(1-p)^{n-k}$. Por exemplo, uma tal seqüência pode ser representada por

$$A_1 \cap \ldots \cap A_k \cap A_{k+1}^c \ldots A_n^c \tag{4.38}$$

onde A_i é o evento em que A ocorre na i-ésima realização do experimento, $i = 1, \ldots, n$. Então a probabilidade desta seqüência vale

$$\begin{aligned} \mathbb{P}(A_1 \cap \ldots \cap A_k \cap A_{k+1}^c \ldots A_n^c) &= \mathbb{P}(A_1) \ldots \mathbb{P}(A_k)\mathbb{P}(A_{k+1}^c) \ldots \mathbb{P}(A_n^c) \\ &= p^k(1-p)^{n-k}, \end{aligned} \tag{4.39}$$

onde a primeira igualdade segue da independência entre as realizações, e a segunda da igualdade de condições entre elas. Neste caso A ocorreu nas k primeiras realizações do experimento, e A^c ocorreu nas $n - k$ realizações seguintes. Uma seqüência genérica pode ser representada como em (4.38), trocando as posições em que os $(n-k)$ sinais de complementar $({}^c)$ aparecem. Nenhuma tal troca modifica a probabilidade da seqüência, devido à fatoração devida à independência, e à igualdade de condições, que faz que $\mathbb{P}(A_i) = p$ para todo i.

O total de seqüências pode então ser identificado como o total de escolhas distintas de $n - k$ posições em $\{1, \ldots, n\}$ para colocar os sinais de complementar, ou equivalentemente, o total de escolhas distintas de k posições em $\{1, \ldots, n\}$ onde *não* colocar os sinais de complementar. Mas isto é o número de escolhas distintas de k elementos num conjunto de tamanho n, que como se sabe é a combinação de n, k a k; em símbolos

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

Concluimos que

$$\mathbb{P}(X = k) = \binom{n}{k} p^k(1-p)^{n-k}, \quad k = 0, 1, \ldots, n. \tag{4.40}$$

Se uma v.a. X tiver distribuição de probabilidades dada por (4.40), dizemos que X tem distribuição *binomial* com os parâmetros n e p, e denotamos $X \sim \text{Bin}(n, p)$.

Observação 4.18 *O experimento aleatório que consiste em observar, em cada realização do experimento aleatório original, se o evento A ocorre ou não é muitas vezes denominado* ensaio de Bernoulli *ou* ensaio binomial*, pois há apenas dois resultados possíveis: A ocorre, e neste caso dizemos que houve* sucesso*, ou A^c ocorre, e neste caso dizemos que houve* fracasso*. Nestes termos, se X é o número de sucessos em n ensaios de Bernoulli independentes, em que a probabilidade de sucesso em cada ensaio vale p, então $X \sim Bin(n, p)$.*

Proposição 4.19 *Se para dados $n \geq 1$ e $p \in [0, 1]$, $X \sim Bin(n, p)$, então*

$$\mathbb{E}(X) = np, \tag{4.41}$$

$$\mathbb{V}(X) = np(1 - p). \tag{4.42}$$

Veremos demonstrações de (4.41) e (4.42) adiante, na Seção 4.4.

Exemplo 4.20 *Em 10 lançamentos de uma moeda honesta, qual é a probabilidade de observarmos*

1. *exatamente 5 caras?*
2. *entre 3 e 7 caras?*
3. *mais do que 7 caras?*

Para responder a estas questões, vamos considerar a v.a. X = "número de caras nos 10 lançamentos da moeda". Supondo que os lançamentos são independentes, e como se trata sempre da mesma moeda, que é honesta, temos que $X \sim Bin(10, 1/2)$ (os lançamentos da moeda são ensaios de Bernoulli em que sucesso é sair cara, cuja probabilidade é sempre $1/2$).

De (4.40),

$$\begin{aligned}\mathbb{P}(X = 5) &= \binom{10}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^5 = \frac{10!}{5!\,5!}\left(\frac{1}{2}\right)^{10} \\ &= \frac{10 \times 9 \times 8 \times 7 \times 6}{5 \times 4 \times 3 \times 2}\frac{1}{1024} = \frac{252}{1024} = 0.246, \end{aligned} \tag{4.43}$$

que é a resposta à primeira pergunta. Para responder às demais perguntas, precisamos calcular $\mathbb{P}(X = k)$ *para* $k \geq 3$. *Vamos apresentar estes cálculos, alguns mais, outros menos explicitamente.*

$$\begin{aligned}
\mathbb{P}(X = 3) &= \binom{10}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^7 = \frac{10!}{3!\,7!}\left(\frac{1}{2}\right)^{10} \\
&= \frac{10 \times 9 \times 8}{3 \times 2}\frac{1}{1024} = \frac{120}{1024} = 0.117, && (4.44) \\
\mathbb{P}(X = 4) &= \binom{10}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^6 = \frac{10!}{4!\,6!}\left(\frac{1}{2}\right)^{10} \\
&= \frac{10 \times 9 \times 8 \times 7}{4 \times 3 \times 2}\frac{1}{1024} = \frac{210}{1024} = 0.205, && (4.45) \\
\mathbb{P}(X = 6) &= \binom{10}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^4 = 0.205, && (4.46) \\
\mathbb{P}(X = 7) &= \binom{10}{7}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^3 = 0.117, && (4.47) \\
\mathbb{P}(X = 8) &= \binom{10}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2 = 0.044, && (4.48) \\
\mathbb{P}(X = 9) &= \binom{10}{9}\left(\frac{1}{2}\right)^9\left(\frac{1}{2}\right)^1 = 0.010, && (4.49) \\
\mathbb{P}(X = 10) &= \binom{10}{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^0 = 0.001 && (4.50)
\end{aligned}$$

Respondendo à segunda pergunta,

$$\begin{aligned}
& \mathbb{P}(3 \leq X \leq 7) \\
= \; & \mathbb{P}(X = 3) + \mathbb{P}(X = 4) + \mathbb{P}(X = 5) + \mathbb{P}(X = 6) + \mathbb{P}(X = 7) \\
= \; & 0.117 + 0.205 + 0.246 + 0.205 + 0.117 = 0.880; && (4.51)
\end{aligned}$$

e à terceira,

$$\begin{aligned}
\mathbb{P}(X \geq 8) &= \mathbb{P}(X = 8) + \mathbb{P}(X = 9) + \mathbb{P}(X = 10) \\
&= 0.044 + 0.010 + 0.001 = 0.056. && (4.52)
\end{aligned}$$

Observação 4.21 *A última probabilidade acima (em (4.52)) é relevante em inferência estatística da seguinte forma. Suponha que não conheçamos a* *probabilidade de cara da moeda, desconfiemos que ela não seja honesta, com um viés para cara, mas queiramos ser cautelosos em rejeitar a hipótese de honestidade. Vamos então procurar reunir* evidências estatísticas *contra a hipótese de honestidade, e medir sua significância.*

Os 10 lançamentos da moeda são então um procedimento de reunião de evidências contrárias à hipótese de honestidade. Suponha que obtenhamos 8 caras: esta é a evidência contra a hipótese de honestidade. Qual sua significância?

A probabilidade em (4.52) é uma medida da significância desta evidência, no seguinte sentido. Se a moeda fosse honesta, qual seria a probabilidade de obtermos 8 caras ou resultado mais significante contra a hipótese de honestidade, na direção de nossa desconfiança de viés para cara? Isto se traduz em $X \geq 8$. *Como sob a hipótese de honestidade* $X \sim Bin(10, 1/2)$, *temos que a probabilidade desejada é a dada em (4.52). Esta probabilidade neste contexto é chamada de* p-valor *associado ao resultado dos lançamentos. Quanto menor o p-valor, maior é a evidência contra a hipótese de honestidade.*

A questão de decidir se a evidência medida por (4.52) é suficientemente forte para finalmente rejeitarmos a hipótese de honestidade é em princípio subjetiva, mas em muitas situações práticas se adota preliminarmente um limiar, *como* 0.01, *ou* 0.05, *ou* 0.10. *Se o p-valor estiver abaixo do limiar, então rejeitamos a hipótese de honestidade; se estiver acima, não a rejeitamos.*

Na situação acima, se adotássemos (preliminarmente) o limiar de 0.05, *então, como o p-valor de* 0.056 *está acima do limiar, não rejeitaríamos a hipótese de honestidade.*

Um resultado de 9 *caras, por outro lado, leva a um p-valor de* 0.011, *e neste caso, com base no limiar adotado, rejeitaríamos a hipótese de honestidade.*

Observação 4.22 *Uma aplicação à amostragem é a seguinte. Suponha que estejamos interessados em conhecer a proporção* p *de indivíduos de certa população com certa característica (por exemplo, uma característica física, ou social, ou de opinião). Se colhermos uma amostra casual simples de* n *indivíduos desta população com reposição, então o número* X *de indivíduos com a característica de interesse na amostra é uma v.a. relevante. Notemos que neste contexto, a observação de cada indivíduo da amostra é um ensaio de Bernoulli (o indivíduo exibe ou não a característica de interesse) independente dos demais (em particular por causa da reposição), e se identificarmos* *sucesso com o indivíduo exibir a característica de interesse, então a probabilidade de sucesso em cada ensaio é p (por causa da reposição). Concluímos que $X \sim Bin(n, p)$.*

Na situação do Exemplo 3.9, se a característica de interesse for sexo feminino, então $X \sim Bin(5, 0.55)$. Podemos por exemplo expressar os eventos A, B e C naquele exemplo em termos de X como segue, e usar a distribuição binomial para calcular suas probabilidades (trabalho que deixamos para o leitor).

$$A = \{X = 0\},\ B = \{X = 3\},\ C = \{X \geq 3\}. \tag{4.53}$$

4.2.4 O modelo hipergeométrico

Voltando à Observação 4.22, suponha que a amostragem seja feita sem reposição. Neste caso perde-se a independência e a igualdade de condições entre os ensaios de Bernoulli, a e v.a. X = "número de sucessos nos n ensaios" deixa de ser binomial.

Vamos então calcular a distribuição de X neste caso. Suponha que $M \geq 2$ seja o tamanho da população (estamos no contexto do parágrafo sobre amostragem sem reposição, na página 94, e usando aquele espaço de probabilidades), e K seja o número de indivíduos da população com a característica de interesse; $n \leq M$ é o tamanho da amostra. Então o conjunto de valores possíveis de X são os números inteiros não negativos entre $0 \vee (n - M + K)$ e $n \wedge K$ (em outras palavras, $\mathcal{V}_X = [0 \vee (n - M + K), n \wedge K] \cap \mathbb{Z}$).

Então, para $k \in \mathcal{V}_X$, o evento $\{X = k\}$ consiste de amostras com k indivíduos dentre os K da população com a característica de interesse, e $n - k$ indivíduos dentre os $M - K$ da população sem tal característica.

O número de tais amostras é pois o número de escolhas de k indivíduos dentre K multiplicado pelo número de escolhas de $n - k$ indivíduos dentre $M - K$. Concluímos que

$$\mathbb{P}(X = k) = \frac{\#\{X = k\}}{\#\Omega} = \frac{\binom{K}{k}\binom{M-K}{n-k}}{\binom{M}{n}}, \quad k \in [0 \vee (n - M + K), n \wedge K] \cap \mathbb{Z}. \tag{4.54}$$

Dizemos então que X tem distribuição *hipergeométrica*, com a notação $X \sim \mathrm{HG}(M, K; n)$.

Pode-se mostrar (mas não o faremos nestas notas) que

$$\mathbb{E}(X) = np, \tag{4.55}$$

$$\mathbb{V}(X) = fnp(1-p), \tag{4.56}$$

onde $p = K/M$ e $f = 1 - (n-1)/(M-1)$. Note que a média coincide com a da distribuição do caso com reposição (em que X é binomial), e que a variância difere da daquele caso pelo fator f (a *fração amostral*).

Na situação do Exemplo 3.10, se a característica de interesse for sexo feminino (como no caso discutido na Observação 4.22, então $X \sim \mathrm{HG}(100, 55; 5)$. Os eventos A, B e C podem ser expressados em termos de X como em (4.53), e suas probabilidades obtidas de (4.54). Deixamos os detalhes para o leitor. No Exemplo 3.13, $X \sim \mathrm{HG}(N, K; n)$.

4.2.5 O modelo geométrico

Suponha que uma moeda cuja probabilidade de cara é $p \in (0, 1]$ seja lançada sucessivamente de maneira independente. Seja X o número de coroas até a primeira cara. Então X é uma v.a. com $\mathcal{V}_X = \{0, 1, 2, \ldots\}$. Para $k \geq 1$, temos que $X = k$ se e somente se (sse) ocorre coroa nos k primeiros lançamentos e cara no k-ésimo lançamento. Logo $\mathbb{P}(X = k) = (1-p)^k p$. Como $X = 0$ sse sai cara no primeiro lançamento, e isto tem probabilidade p, temos que

$$\mathbb{P}(X = k) = (1-p)^k p, \ k \in \mathcal{V}_X. \tag{4.57}$$

Neste caso, dizemos que X tem distribuição *geométrica* com parâmetro p, com a notação $X \sim \mathrm{G}(p)$.

Proposição 4.23 *Se $X \sim G(p)$, então*

$$\mathbb{P}(X > k) = (1-p)^{k+1}, \ k \geq 0, \tag{4.58}$$

$$\mathbb{E}(X) = \frac{1-p}{p}, \tag{4.59}$$

$$\mathbb{V}(X) = \frac{1-p}{p^2}. \tag{4.60}$$

Demonstração Para $k \geq 1$, $X > k$ sse ocorre coroa nos $k+1$ primeiros lançamentos. (4.58) segue. De (4.16) e (4.58),

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} (1-p)^{k+1} = \frac{1-p}{1-(1-p)} = \frac{1-p}{p}. \tag{4.61}$$

De (4.17) e (4.58),

$$\mathbb{E}(X^2) = \sum_{k=0}^{\infty} k\,(1-p)^{k+1} + \frac{1-p}{p}. \tag{4.62}$$

Mas

$$\sum_{k=0}^{\infty} k\,(1-p)^{k+1} = \frac{1-p}{p}\sum_{k=1}^{\infty} k\,(1-p)^k p = \frac{1-p}{p}\,\mathbb{E}(X) = \left(\frac{1-p}{p}\right)^2. \tag{4.63}$$

Substituindo (4.63) em (4.62), vem

$$\mathbb{E}(X^2) = 2\left(\frac{1-p}{p}\right)^2 + \frac{1-p}{p}. \tag{4.64}$$

De (4.24), (4.59) e (4.64), segue (4.60).

No Exemplo 3.5 acima, vimos o caso $p = 1/2$.

Observação 4.24 *Na literatura a distribuição geométrica é muitas vezes identificada com a v.a.* X' = *número de lançamentos da moeda até sair a primeira cara. A relação com* X *acima é então* $X' = X + 1$.

4.2.6 O modelo binomial negativo

No contexto da subseção acima, suponha que X seja o número de coroas até sair a n-ésima cara, onde $n \geq 1$ é um parâmetro. Então $\mathcal{V}_X = \{0, 1, 2, \ldots\}$ e para $k \geq 1$, $X = k$ sse o $(n+k)$-ésimo lançamento resulta em cara, e nos $n+k-1$ lançamentos anteriores, ocorrem $n-1$ caras e k coroas em qualquer ordem. Logo

$$\mathbb{P}(X = k) = \binom{n+k-1}{k} p^n (1-p)^k,\ k \geq 0. \tag{4.65}$$

(Note que a fórmula funciona no caso $k = 0$ também.)

Se uma v.a. X tiver distribuição de probabilidades dada por (4.65), dizemos que X tem distribuição *binomial negativa* com os parâmetros n e p, e denotamos $X \sim \mathrm{BN}(n, p)$.

Proposição 4.25 *Se para dados $n \geq 1$ e $p \in (0,1]$, $X \sim BN(n,p)$, então*

$$\begin{aligned} \mathbb{E}(X) &= n\frac{1-p}{p}, & (4.66) \\ \mathbb{V}(X) &= n\frac{1-p}{p^2}. & (4.67) \end{aligned}$$

Veremos demonstrações de (4.66) e (4.67) adiante, na Seção 4.4.

4.2.7 O modelo de Poisson

Suponha que estejamos interessados num evento raro de um experimento aleatório. Um evento raro é um evento de probabilidade pequena. Logo, ele provavelmente não será observado em uma realização do experimento. Será necessário repetir o experimento (de forma independente) um bom número de vezes para termos uma chance razoável de observar tal evento.

Para precisar um pouco mais a situação, vamos imaginar que A é o evento, e $p = \mathbb{P}(A)$ sua probabilidade. Vamos imaginar que p é bastante próximo de 0. Sendo X o número de ocorrências de A em n repetições independentes do experimento, $X \sim \text{Bin}(n,p)$, teríamos de ter n da ordem de $1/p$ para termos uma chance razoável de observar A pelo menos uma vez (para que o valor esperado $\mathbb{E}(X) = np$ fique da ordem de 1).

Vamos inverter o raciocínio e imaginar que $p = \lambda/n$, onde λ é um parâmetro positivo, e n é um número inteiro bastante grande (de forma que p seja bastante pequeno). Então $\mathbb{E}(X) = \lambda$. Temos ainda de (4.40) que para $k \geq 0$

$$\begin{aligned} \mathbb{P}(X = k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \left[\frac{n!}{(n-k)!}\frac{1}{n^k}\right]\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}. \qquad (4.68) \end{aligned}$$

Como estamos pensando em n bastante grande, vamos tomar o limite da expressão à direita de (4.68) quando $n \to \infty$. A expressão entre colchetes pode ser reescrita da seguinte forma.

$$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)$$

Como k está fixo, o limite da expressão acima quando $n \to \infty$ é o produto do limite de cada fator, que vale 1. Logo o limite da expressão também vale 1. Pelo mesmo motivo,

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = 1.$$

Resta avaliar o limite de

$$\left(1-\frac{\lambda}{n}\right)^{n}$$

quando $n \to \infty$. Este sabidamente vale $e^{-\lambda}$. Concluímos que o limite da expressão à direita de (4.68) quando $n \to \infty$ vale

$$e^{-\lambda}\,\frac{\lambda^k}{k!}. \tag{4.69}$$

Sabe-se também que

$$\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{\lambda}, \tag{4.70}$$

logo as expressões em (4.69), $k = 0, 1, 2, \ldots$ são a distribuição de probabilidade de uma v.a. Y com $\mathcal{V}_Y = \mathbb{N}$.

$$\mathbb{P}(Y = k) = e^{-\lambda}\,\frac{\lambda^k}{k!},\ \ k = 0, 1, 2, \ldots \tag{4.71}$$

e neste caso dizemos que Y tem distribuição *de Poisson* com parâmetro λ, e denotamos $Y \sim \text{Po}(\lambda)$.

Em virtude da discussão acima, dizemos que a distribuição de Poisson aproxima a distribuição binomial quando p é pequeno e n é grande (de forma que np não seja nem grande nem pequeno).

Exemplo 4.26 *Suponha que os erros tipográficos na edição de um livro de 300 páginas sejam distribuídos aleatoriamente com uma média de 1 página com erros a cada* 100*. Qual a probabilidade de acharmos no livro todo*

1. *nenhuma página com erros tipográficos?*

2. *pelo menos 2 páginas com erros tipográficos?*

3. *entre 1 e 3 páginas com erros tipográficos?*

Para responder a estas questões, vamos considerar a variável Y = número de páginas com erros tipográficos no livro. Vamos supor que a probabilidade de erro numa página é $1/100 = 3/300$. Neste caso podemos pelo menos de forma aproximada dizer que $Y \sim Po(3)$. Então, de (4.71)

1. $\mathbb{P}(Y = 0) = e^{-3} = 0.05$
2. $\mathbb{P}(Y \geq 2) = 1 - \mathbb{P}(Y = 0) - \mathbb{P}(Y = 1) = 1 - e^{-3} - 3e^{-3} = 0.80$
3. $\mathbb{P}(1 \leq Y \leq 3) = \mathbb{P}(Y = 1) + \mathbb{P}(Y = 2) + \mathbb{P}(Y = 3)$
   $= 3e^{-3} + \frac{3^2}{2}e^{-3} + \frac{3^3}{6}e^{-3} = 0.60.$

Proposição 4.27 *Suponha que $Y \sim Po(\lambda)$, então*

$$\mathbb{E}(Y) = \mathbb{V}(Y) = \lambda. \tag{4.72}$$

Demonstração Não faremos um argumento direto, mas usaremos a aproximação binomial discutida acima. Como a distribuição de Y é o limite da de $X \sim \mathrm{Bin}(n, \lambda/n)$ quando $n \to \infty$, é natural supormos que

$$\mathbb{E}(Y) = \lim_{n\to\infty} \mathbb{E}(X) = \lim_{n\to\infty} n\frac{\lambda}{n} = \lambda, \tag{4.73}$$

$$\mathbb{V}(Y) = \lim_{n\to\infty} \mathbb{E}(X) = \lim_{n\to\infty} n\frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right) = \lambda. \tag{4.74}$$

Esta suposição pode ser justificada, e temos (4.72).

4.3 Modelos para variáveis aleatórias contínuas

Nesta seção apresentamos alguns modelos para v.a.'s contínuas. O principal deles, o modelo normal, já foi visto como modelo de distribuição de freqüencias para variáveis populacionais.

4.3.1 O modelo uniforme

Dado um intervalo finito $[a, b]$ da reta, dizemos que uma v.a. X tem distribuição *uniforme* em $[a, b]$, com a notação $X \sim \mathrm{U}([a, b])$, se $\mathcal{V}_X = [a, b]$ e a função densidade de probabilidade de X for

$$f_X = \frac{1}{b-a}\, 1_{[a,b]}. \tag{4.75}$$

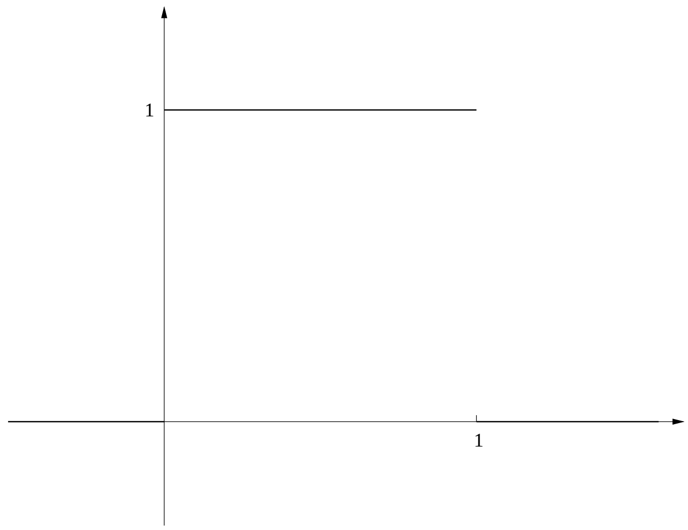

Figura 4.3

O gráfico de f_X no caso $a = 0$, $b = 1$ é apresentado na Figura 4.3. No Exemplo 4.3, temos $X \sim \mathrm{U}([0, 1])$.

Proposição 4.28 *Suponha que* $X \sim U([a, b])$, *então*

$$\begin{aligned}
\mathbb{E}(X) &= \frac{a+b}{2}, && (4.76)\\
\mathbb{V}(X) &= \frac{(b-a)^2}{12}. && (4.77)
\end{aligned}$$

Demonstração De (4.6) e (4.75) temos

$$\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{\infty} x \, \frac{1}{b-a} \, 1_{[a,b]}(x) \, dx = \frac{1}{b-a} \int_a^b x \, dx \\
&= \frac{1}{b-a} \, \frac{b^2 - a^2}{2} = \frac{a+b}{2}, && (4.78)
\end{aligned}$$

e temos (4.76). De forma similar

$$\begin{aligned}
\mathbb{E}(X^2) &= \int_{-\infty}^{\infty} x^2 \, \frac{1}{b-a} \, 1_{[a,b]}(x) \, dx = \frac{1}{b-a} \int_a^b x^2 \, dx \\
&= \frac{1}{b-a} \, \frac{b^3 - a^3}{3} = \frac{a^2 + ab + b^2}{3}, && (4.79)
\end{aligned}$$

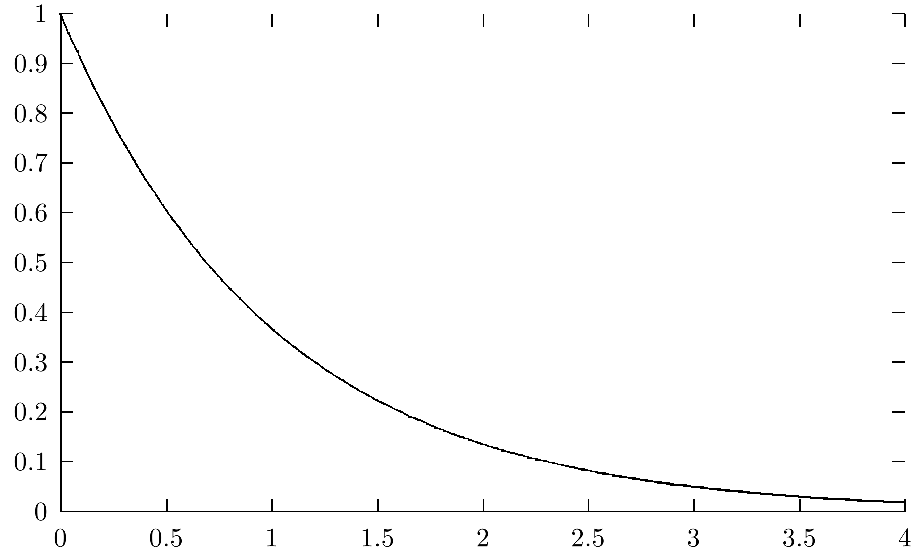

Figura 4.4

e de (4.24)

$$\mathbb{V}(X) = \frac{4(a^2 + ab + b^2) - 3(a+b)^2}{12} = \frac{(b-a)^2}{12}, \tag{4.80}$$

e temos (4.77).

4.3.2 O modelo exponencial

Dado $\lambda > 0$, dizemos que uma v.a. X tem distribuição *exponencial* com parâmetro λ, com a notação $X \sim \text{Exp}(\lambda)$, se $\mathcal{V}_X = (0, \infty)$ e a função densidade de probabilidade de X for

$$f_X(x) = \lambda\, e^{-\lambda x}\, 1_{(0,\infty)}(x). \tag{4.81}$$

O gráfico de f_X no caso $\lambda = 1$ é apresentado na Figura 4.4.

Proposição 4.29 *Se $X \sim Exp(\lambda)$, então*

$$\begin{aligned} \mathbb{P}(X > x) &= e^{-\lambda x},\ x > 0 & (4.82)\\ \mathbb{E}(X) &= \frac{1}{\lambda}, & (4.83)\\ \mathbb{V}(X) &= \frac{1}{\lambda^2}. & (4.84) \end{aligned}$$

Demonstração

$$\mathbb{P}(X > x) = \int_x^\infty f_X(y)\,dy = \int_x^\infty \lambda\, e^{-\lambda\, x}\,dx = \lambda \int_x^\infty e^{-\lambda\, x}\,dx, \qquad (4.85)$$

e (4.82) segue de (A.6) com $a = x$ e $b = \infty$ (neste caso $e^{-\lambda\,\infty} = 0$).

De (4.18),

$$\mathbb{E}(X) = \int_0^\infty e^{-\lambda\, x}\,dx \overset{(A.6)}{=} \frac{1}{\lambda}, \qquad (4.86)$$

e temos (4.83).

De (4.19),

$$\mathbb{E}(X^2) = 2\int_0^\infty x\, e^{-\lambda\, x}\,dx = \frac{2}{\lambda}\int_0^\infty x\,\lambda\, e^{-\lambda\, x}\,dx = \frac{2}{\lambda}\,\mathbb{E}(X) = \frac{2}{\lambda^2}, \qquad (4.87)$$

e (4.84) segue de (4.24), (4.87) e (4.83).□

A distribuição exponencial é muito usada como modelo para tempos de espera entre eventos, tempos de vida, e em situação semelhantes.

Exemplo 4.30 *O tempo de vida de uma lâmpada é uma v.a.* T *com distribuição exponencial de parâmetro* $1/2$*. Qual é a probabilidade de a lâmpada durar*

1. *menos do que* 2 *unidades de tempo?*
2. *pelo menos 3 unidades detempo?*
3. *entre 1 e 3 unidades de tempo?*

1. $\mathbb{P}(T < 2) = \frac{1}{2}\int_0^2 e^{-x/2}\,dx = \frac{1}{2}\int_0^2 e^{-x/2}\,dx \overset{(A.6)}{=} 1 - e^{-1}$
2. $\mathbb{P}(T \geq 3) = \frac{1}{2}\int_3^\infty e^{-x/2}\,dx = \frac{1}{2}\int_3^\infty e^{-x/2}\,dx \overset{(A.6)}{=} e^{-1.5}$
3. $\mathbb{P}(1 \leq T \leq 3) = \frac{1}{2}\int_1^3 e^{-x/2}\,dx = \frac{1}{2}\int_1^3 e^{-x/2}\,dx \overset{(A.6)}{=} e^{-0.5} - e^{-1.5}$

Observação 4.31 (Falta de memória da distribuição exponencial) *Suponha que* $X \sim$ *Exp*(λ) *e que saibamos que* $X > x$ *para algum* $x \geq 0$*. Vamos calcular a probabilidade de que* $X > x + y$ *para algum* $y \geq 0$*.*

$$\mathbb{P}(X > x + y | X > x) = \frac{\mathbb{P}(\{X > x + y\} \cap \{X > x\})}{\mathbb{P}(X > x)}. \qquad (4.88)$$

Mas como $y > 0$, temos que $\{X > x + y\} \cap \{X > x\} = \{X > x + y\}$, então

$$\mathbb{P}(X > x + y|X > x) = \frac{\mathbb{P}(X > x + y)}{\mathbb{P}(X > x)} \overset{(4.82)}{=} \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}} = e^{-\lambda y} \overset{(4.82)}{=} \mathbb{P}(X > y). \tag{4.89}$$

E temos que

$$\mathbb{P}(X > x + y|X > x) = \mathbb{P}(X > y) \quad \textit{para } x, y \geq 0. \tag{4.90}$$

Logo, se $X > x$, o valor excedente de X além de x tem distribuição independente de x, e logo igual ao caso $x = 0$, em que o excedente é a própria X. Esta propriedade é por isto denominada falta de memória da distribuição exponencial.

Exemplo 4.32 *Suponha que a lâmpada do Exemplo 4.30 acima esteja ativada numa sala fechada. Ao entrar na sala, você nota que ela está acesa. Qual a probabilidade de ela durar ainda por 2 unidades de tempo? Por enquanto tempo você espera que ela ainda dure?*

Pela falta de memória da distribuição exponencial, não importa por quanto tempo t a lâmpada tenha estado acesa quando você entra na sala, o fato de ela estar acesa diz que $X > t$, e dado isto, pela falta de memória, o tempo adicional de funcionamento continua Exp(2). Então as respostas às perguntas acima são

$$\mathbb{P}(X > t + 2|X > t) \overset{(4.90)}{=} \mathbb{P}(X > 2) \overset{(4.82)}{=} e^{-1}, \tag{4.91}$$

$$\begin{aligned} \mathbb{E}(X - t|X > t) &\overset{(4.18)}{=} \int_0^\infty \mathbb{P}(X - t > x|X > t)\, dx \\ &= \int_0^\infty \mathbb{P}(X > t + x|X > t)\, dx \\ &\overset{(4.90)}{=} \int_0^\infty \mathbb{P}(X > x)\, dx \overset{(4.18)}{=} \mathbb{E}(X) = 2. \end{aligned} \tag{4.92}$$

4.3.3 O modelo normal

Dados números $\mu \in \mathbb{R}$ e $\sigma > 0$, uma v.a. contínua X é dita ter distribuição normal com os parâmetros μ e σ^2 se $\mathcal{V}_X = \mathbb{R}$ e f_X for dada por (1.20), com a mesma notação do caso de X ser uma variável populacional com distribuição normal com parâmetros μ e σ^2: $X \sim \mathrm{N}(\mu, \sigma^2)$. (Veja a Subseção 1.2.1.)

Claramente, esta distribuição de probabilidades tem as mesmas propriedades matemáticas da distribuição de freqüências estudada na Subseção 1.2.1. Em particular

$$\begin{aligned}\mathbb{E}(X) &= \mu, && (4.93)\\ \mathbb{V}(X) &= \sigma^2. && (4.94)\end{aligned}$$

(Veja (2.13) e (2.77).)

A maneira de calcular probabilidades envolvendo variável normal, fazendo padronização, e usando simetria e complementaridade, funciona exatamente como na Subseção 1.2.1.

A título de exemplo, suponha que X seja uma v.a. tal que $X \sim \mathrm{N}(48, 625)$. Esta é a mesma distribuição daquela do Exemplo 1.8. As freqüências ali calculadas podem então ser vistas como probabilidades no presente caso.

Aproximação normal para a distribuição binomial

Uma das propriedades importantes da distribuição normal (vista como freqüência ou probabilidade) é que ela aproxima outras distribuições, tornando simples o cálculo de probabilidades ou freqüências (de forma aproximada) em casos complicados. Vamos apresentar neste parágrafo o caso da aproximação normal para a distribuição binomial.

Seja $X \sim \mathrm{Bin}(16, 0.5)$. Na Figura 4.5 representamos a distribuição de X com barras de bases centradas nos valores de $\mathcal{V}_X$ e cujas alturas são as respectivas probabilidades. Note que as bases das barras têm comprimento 1, logo as probabilidades são também as áreas das barras. Podemos então pensar no gráfico de barras de X como histograma de X. A silhueta deste gráfico é reminiscente da curva normal. Vamos então ajustar uma curva normal neste gráfico. Mas qual curva? (Com quais parâmetros?) É natural tomar para a distribuição normal as mesmas média e variância da distribuição binomial. Vamos então considerar $Y \sim \mathrm{N}(8, 4)$ (isto é com $\mu = \mathbb{E}(X) = 16 \times 0.5$ e $\sigma^2 = \mathbb{V}(X) = 16 \times 0.5 \times 0.5$). Na Figura 4.6 superpomos o gráfico de barras de X e o gráfico da densidade de Y para realçar a similaridade. Consideremos agora a seguinte probabilidade binomial.

$$\mathbb{P}(8 \leq X \leq 12) \qquad (4.95)$$

Em termos do gráfico de barras de X, conforme argumentamos acima, esta probabilidade é a soma das áreas das barras centradas em 8, 9, ..., 12; em outras palavras, a área no histograma de barras entre 7.5 e 12.5 (pois a base

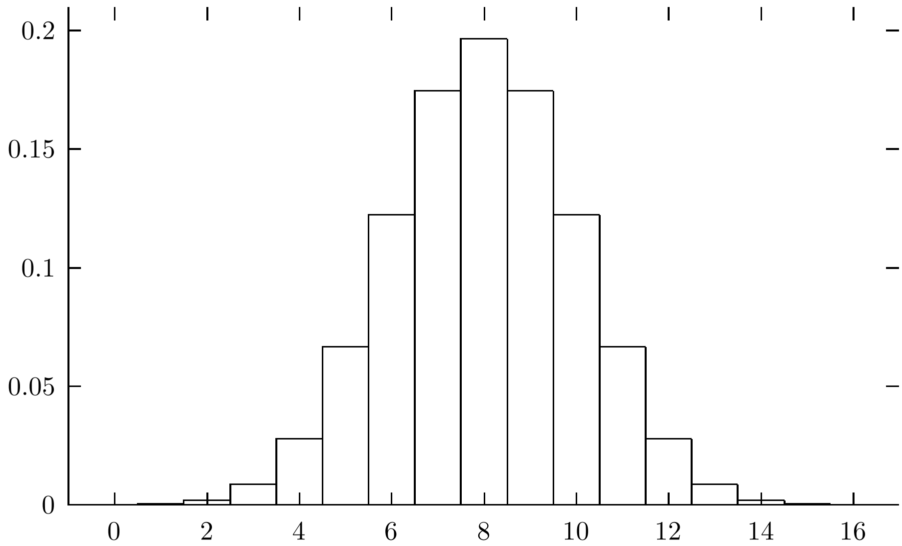

Figura 4.5 Gráfico de barras de X

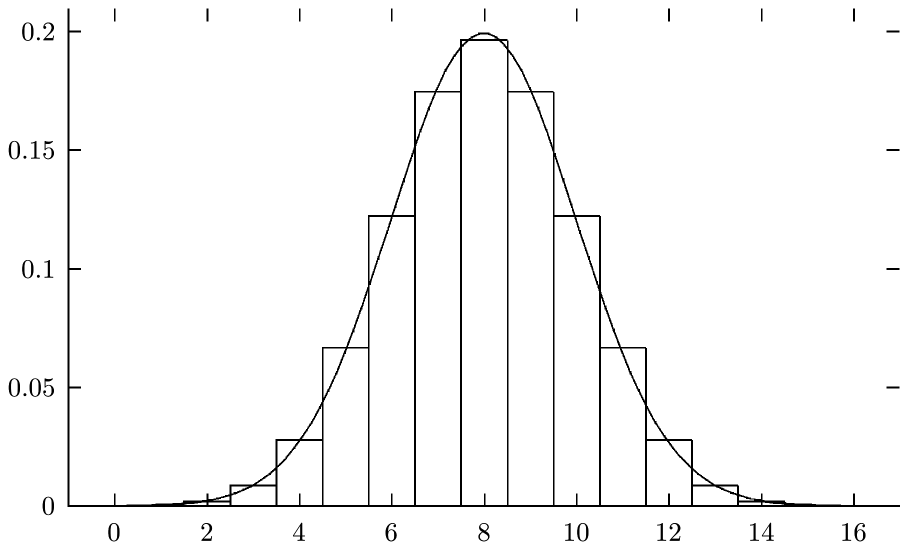

Figura 4.6 Gráfico de barras de X com densidade de Y superposta.

de cada barra começa em $k - 0.5$ e termina em $k + 0.5$, onde k é o ponto central da base). É claro da Figura 4.6 que a área sob o histograma normal de Y é uma boa aproximação para a probabilidade em questão. De fato, de (4.40), achamos

$$\mathbb{P}(8 \leq X \leq 12) = 0.587555; \tag{4.96}$$

e da aproximação normal, temos

$$\begin{aligned} \mathbb{P}(8 \leq X \leq 12) &\approx \mathbb{P}(7.5 \leq Y \leq 12.5) = \mathbb{P}(-0.25 \leq Z \leq 2.25) \\ &= A(2.25) + A(0.25) - 1 = 0.987776 + 0.598706 - 1 \\ &= 0.586482. \end{aligned} \tag{4.97}$$

Em geral, para $X \sim \mathrm{Bin}(n, p)$, a v.a. normal adequada para aproximar X é $Y \sim \mathrm{N}(np, np(1-p))$, isto é, a v.a. normal com média $\mu = \mathbb{E}(X) = np$ e variância $\sigma^2 = \mathbb{V}(X) = np(1-p)$. Neste caso, temos, para $0 \leq i \leq j \leq n$

$$\mathbb{P}(i \leq X \leq j) \approx \mathbb{P}(i - 0.5 \leq Y \leq j + 0.5). \tag{4.98}$$

A aproximação é tanto melhor quanto

1. mais central for p (isto é, mais próximo de 0.5);
2. maior for n;
3. um critério que combina os itens anteriores: quanto maior for $np(1-p)$.

Algumas referências dão: $np(1-p) \geq 3$ implica em boa aproximação.

Observação 4.33 *Uma aproximação mais grosseira do que (4.98), mas bastante usada na prática é*

$$\mathbb{P}(i \leq X \leq j) \approx \mathbb{P}(i \leq Y \leq j). \tag{4.99}$$

Note que em (4.99) descartamos metade da área de cada uma das barras centradas em i e j.

Observação 4.34 *Uma afirmação mais precisa sobre a aproximação normal é a dada pelo enunciado do* Teorema Central do Limite*: seja $X_n \sim Bin(n, p)$, $p \in (0, 1)$, e seja Z_n a v.a. X_n padronizada:*

$$Z_n = \frac{X_n - \mathbb{E}(X_n)}{\mathbb{DP}(X_n)} = \frac{X_n - np}{\sqrt{np(1-p)}}.$$

Então a distribuição de Z_n converge para a distribuição normal padrão quando $n \to \infty$, isto é, para todo $a, b \in \mathbb{R}$ com $a < b$:

$$\lim_{n\to\infty} \mathbb{P}(a < Z_n < b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx. \tag{4.100}$$

Exemplo 4.35 *Num referendo a ser realizado em dado local,* 55% *da população é pelo "Sim" e* 45% *pelo "Não". Planeja-se uma pesquisa de opinião a respeito com base numa amostra casual simples com reposição de tamanho* 100 *da população. Qual a probabilidade de a maioria pelo "Sim" não aparecer na amostra?*

Seja X o número de indivíduos pelo "Sim" na amostra. Então $X \sim Bin(100, 0.55)$. Queremos determinar $\mathbb{P}(X \le 50)$. De (4.98), sendo $Y \sim N(55, 24.75)$, temos

$$\mathbb{P}(X \le 50) \approx \mathbb{P}(Y \le 50.5) = \mathbb{P}(Z \le -0.90) = 1 - A(0.90) = 0.184. \tag{4.101}$$

E se $n = 400$? Neste caso, queremos $\mathbb{P}(X' \le 200)$, onde $X' \sim Bin(400, 0.55)$. A v.a. normal adequada é $Y' \sim N(220, 99)$, e

$$\mathbb{P}(X' \le 200) \approx \mathbb{P}(Y' \le 200.5) = \mathbb{P}(Z \le -1.96) = 1 - A(1.96) = 0.025. \tag{4.102}$$

4.4 Várias variáveis aleatórias

Podemos estar interessados em mais de uma v.a. de dado experimento aleatório. No Exemplo 3.6, as coordenadas X e Y do ponto escolhido são duas v.a.'s. No Exemplo 3.7, os números lançados nos dois lançamentos também são duas v.a.'s.

Dadas $X_1, X_2, \ldots, X_n$, $n \ge 2$, v.a.'s de um mesmo experimento aleatório, a informação que nos interessa sobre elas é aquela dada pela distribuição de probabilidades conjunta de $X_1, X_2, \ldots$ Vamos considerar apenas o caso discreto.

Suponha que $(\Omega, \mathcal{E}, \mathbb{P})$ seja um modelo probabilístico para um dado experimento aleatório e que $X_1, \ldots, X_n$ sejam v.a.'s deste experimento, isto é, $X_i : \Omega \to \mathbb{R}$, $i = 1, \ldots, n$. Vamos supor ainda que cada v.a. seja discreta, isto é, o conjunto de valores de X_i, $\mathcal{V}_i$, é discreto: finito ou infinito enumerável.

Neste caso, a distribuição de probabilidades conjunta de $X_1, \ldots, X_n$ é dada pela função de probabilidades conjunta de $X_1, \ldots, X_n$ a seguir.

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n);\ x_i \in \mathcal{V}_i,\ i = 1, \ldots, n. \tag{4.103}$$

Exemplo 4.36 *No Exemplo 3.7, seja X o número lançado no primeiro lançamento e Y aquele resultante do segundo lançamento.Então*

$$\mathbb{P}(X = i, Y = j) = \frac{1}{36},\ i, j = 1, \ldots, 6. \tag{4.104}$$

Exemplo 4.37 *No Exemplo 3.18, seja X o indicador de que a primeira bola é azul, isto é, $X = 1_{A_1}$, e Y o indicador de que a segunda bola é azul, isto é, $Y = 1_{A_2}$. Então a distribuição conjunta de (X, Y) é dada por (veja a Figura 3.12)*

$$\begin{aligned}
\mathbb{P}(X = 1, Y = 1) &= \mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1) \\
&= 0.6 \times 0.56 = 0.34, &(4.105)\\
\mathbb{P}(X = 1, Y = 0) &= \mathbb{P}(A_1 \cap A_2^c) = \mathbb{P}(A_1)\mathbb{P}(A_2^c|A_1) \\
&= 0.6 \times 0.44 = 0.26, &(4.106)\\
\mathbb{P}(X = 0, Y = 1) &= \mathbb{P}(A_1^c \cap A_2) = \mathbb{P}(A_1^c)\mathbb{P}(A_2|A_1^c) \\
&= 0.4 \times 0.67 = 0.27, &(4.107)\\
\mathbb{P}(X = 0, Y = 0) &= \mathbb{P}(A_1^c \cap A_2^c) = \mathbb{P}(A_1^c)\mathbb{P}(A_2^c|A_1^c) \\
&= 0.4 \times 0.33 = 0.13. &(4.108)
\end{aligned}$$

Uma maneira de obter várias v.a.'s em amostragem é quando sorteamos um indivíduo numa população e consideramos diversas variáveis populacionais medidas no indivíduo sorteado. No Exemplo 1.1, se tomarmos uma amostra casual simples de tamanho 1 da população de funcionários da companhia, e considerarmos a idade e número de filhos do funcionário assim sorteado, temos então duas v.a.'s aleatórias do sorteio.

Uma outra forma é quando tomamos amostra de mais de um indivíduo e medimos a mesma variável populacional nos diversos indivíduos sortedos. No Exemplo 1.1, poderíamos tomar uma amostra casual simples de tamanho 2 da população. X seria o número de filhos do primeiro indivíduo da amostra e Y o número de filhos do segundo indivíduo.

A distribuição conjunta de probabilidades tem as mesmas propriedades matemáticas que a distribuição conjunta de freqüências. Em particular a

Proposição 1.20 é válida para distribuições conjuntas de probabilidades. A versão da Proposição 2.7 para v.a.'s é seguinte.

Sejam $X_1, \ldots, X_n$ v.a.'s (discretas) de um mesmo espaço de probabilidades $(\Omega, \mathcal{E}, \mathbb{P})$ e $h : \mathbb{R}^n \to \mathbb{R}$. Então $Y = h(X_1, \ldots, X_n)$ é uma v.a. de $(\Omega, \mathcal{E}, \mathbb{P})$ e

$$\mathbb{E}(Y) = \mathbb{E}(h(X_1, \ldots, X_n)) = \sum_{x_1,\ldots,x_n} h(x_1, \ldots, x_n)\, \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n). \tag{4.109}$$

Podemos também obter a propriedade de "linearidade" da esperança (veja (2.35)): para constantes numéricas $a_0, a_1, \ldots, a_n$ arbitrárias,

$$\mathbb{E}(a_0 + a_1 X_1 + \ldots + a_n X_n) = a_0 + a_1 E(X_1) + \ldots + a_n \mathbb{E}(X_n). \tag{4.110}$$

Ambas as propriedades (4.109) e (4.110) são válidas em geral (as v.a's não precisam ser discretas; a primeira propriedade tem uma forma um pouco diferente em geral.). Em particular, se X for uma v.a. contínua com densidade f_X, então

$$\mathbb{E}(h(X)) = \int_{-\infty}^{\infty} h(x)\, f_X(x)\, dx. \tag{4.111}$$

(veja (2.22) e (4.10)).

4.4.1 Condicionamento e independência

Dadas duas v.a.'s (discretas) X, Y num mesmo espaço de probabilidades, dado $y \in \mathcal{V}_Y$, a distribuição condicional de probabilidades de X dado $Y = y$ é a coleção de probabilidades condicionais

$$\mathbb{P}(X = x | Y = y),\ x \in \mathcal{V}_X, \tag{4.112}$$

lembrando que $\mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}$.

As distribuições condicionais de probabilidades são matematicamente idênticas às distribuições condicionais de freqüências, e logo têm as mesmas propriedades matemáticas: veja as Observações 1.23 e 1.24.

A *esperança condicional* de X dado $Y = y$ é a esperança da distribuição condicional de X dado $Y = y$:

$$\mathbb{E}(X|Y = y) = \sum_{x \in \mathcal{V}_X} x\, \mathbb{P}(X = x|Y = y) \tag{4.113}$$

(veja (2.38)), e sendo $h : \mathcal{V}_Y \to \mathbb{R}$ tal que $h(y) = \mathbb{E}(X|Y = y)$, então denotamos $h(Y) = \mathbb{E}(X|Y)$ e temos

$$\mathbb{E}(h(Y)) = \mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X) \tag{4.114}$$

(veja (2.39)).

Exemplo 4.38 *Um experimento é realizado em dois estágios. O primeiro estágio consiste em observar uma variável de Poisson Y com parâmetro $\lambda > 0$. No segundo estágio, dado que $Y = n$ no primeiro estágio, lança-se uma moeda com probabilidade de cara $\alpha \in [0, 1]$ n vezes. Seja X o número de caras observadas nos Y lançamentos.*

Temos então que para $n \geq 0$, $X|Y = n \sim Bin(n, \alpha)$, isto é

$$\mathbb{P}(X = k|Y = n) = \binom{n}{k} \alpha^k (1-\alpha)^{n-k}, \ k = 0, \dots, n. \tag{4.115}$$

(Se $n = 0$, então $X \equiv 0$.)

Vamos calcular a distribuição (marginal) de X. Da propriedade correspondente a (1.74) (veja Observação 1.24), temos que se $k \geq 0$,

$$\begin{aligned} & \mathbb{P}(X = k) \\ = & \sum_{n \geq 0} \mathbb{P}(X = k|Y = n)\mathbb{P}(Y = n) = \sum_{n \geq k} \binom{n}{k} \alpha^k (1-\alpha)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} \\ = & \frac{(\alpha\lambda)^k}{k!} e^{-\lambda} \sum_{n \geq k} \frac{((1-\alpha)\lambda)^{n-k}}{(n-k)!} = \frac{(\alpha\lambda)^k}{k!} e^{-\lambda} \sum_{n \geq 0} \frac{((1-\alpha)\lambda)^n}{n!} \\ = & \frac{(\alpha\lambda)^k}{k!} e^{-\lambda} e^{(1-\alpha)\lambda} = e^{-\alpha\lambda} \frac{(\alpha\lambda)^k}{k!}. \end{aligned} \tag{4.116}$$

Logo

$$X \sim Po(\alpha\lambda). \tag{4.117}$$

Daí podemos concluir que $\mathbb{E}(X) = \alpha\lambda$, mas, se estivéssemos interessados apenas na esperança de X, podíamos dispensar o cálculo em (4.116) e a conclusão em (4.117). Bastaria usar (4.114), como se segue.

Temos que para todo $n \geq 0$, $\mathbb{E}(X|Y = n) = n\alpha = \alpha n$. Logo, $\mathbb{E}(X|Y) = \alpha Y$. De (4.114)

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(\alpha Y) = \alpha\,\mathbb{E}(Y) = \alpha\lambda. \tag{4.118}$$

Independência

De forma análoga ao caso de variáveis populacionais, dadas duas v.a.'s X e Y discretas (no mesmo espaço de probabilidades), dizemos que X é independente de Y se a distribuição condicional de X dado $Y = y$ é igual à distribuição marginal de X para todo $y \in \mathcal{V}_Y$. Em outros palavras, se

$$\mathbb{P}(X = x|Y = y) = \mathbb{P}(X = x), \quad \text{para todo } x \in \mathcal{V}_X,\, y \in \mathcal{V}_Y. \tag{4.119}$$

Esta condição é equivalente a

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y), \quad \text{para todo } x \in \mathcal{V}_X,\, y \in \mathcal{V}_Y, \tag{4.120}$$

e de novo temos a condição simétrica de fatoração da probabilidade conjunta nas probabilidades marginais respectivas. Dizemos também, apoiados por esta simetria, que X e Y são independentes (entre si).

Exemplo 4.39 *No Exemplo 4.36, X e Y são independentes pois, para todo $i, j = 1, \ldots, 6$, temos*

$$\mathbb{P}(X = i, Y = j) = \frac{1}{36} = \frac{1}{6} \times \frac{1}{6} = \mathbb{P}(X = i)\mathbb{P}(Y = j),$$

verificando (4.120).

No caso de mais de duas v.a.'s discretas $X_1, \ldots, X_n$, dizemos que são (coletivamente) independentes se

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \ldots \mathbb{P}(X_n = x_n), \tag{4.121}$$

para todo $x_i \in \mathcal{V}_{X_i}$, $i = 1, \ldots, n$.

Uma maneira de obter v.a.'s independentes em amostragem é quando tomamos uma amostra casual simples de tamanho 1 da população em que temos variáveis populacionais independentes. Estas variáveis medidas no indivíduo sorteado são v.a.'s independentes.

Observação 4.40 *Uma outra forma é quando tomamos uma amostra casual simples* com reposição *de mais de um indivíduo e medimos a mesma variável populacional nos diversos indivíduos sortedos. Como argumentado na Observação 3.22, as v.a.'s resultantes são independentes (além disto, tem cada uma a mesma distribuição idêntica à distribuição (de freqüências) da variável populacional).*

Proposição 4.41 *Suponha que $X_1, \ldots, X_n$ sejam v.a.'s independentes. Então*

$$\mathbb{E}(X_1 \times \ldots \times X_n) = \mathbb{E}(X_1) \times \ldots \times \mathbb{E}(X_n). \tag{4.122}$$

Demonstração Vale em geral, mas argumentaremos apenas o caso discreto. De (4.109)

$$\begin{aligned}
\mathbb{E}(X_1 \ldots X_n) &= \sum_{x_1,\ldots,x_n} x_1 \ldots x_n \, \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \\
&= \sum_{x_1,\ldots,x_n} x_1 \ldots x_n \, \mathbb{P}(X_1 = x_1) \ldots \mathbb{P}(X_n = x_n) \\
&= \sum_{x_1,\ldots,x_n} x_1 \, \mathbb{P}(X_1 = x_1) \ldots x_n \, \mathbb{P}(X_n = x_n) \\
&= \sum_{x_1} x_1 \, \mathbb{P}(X_1 = x_1) \ldots \sum_{x_n} x_n \, \mathbb{P}(X_n = x_n) \\
&= \left(\sum_{x_1} x_1 \, \mathbb{P}(X_1 = x_1) \right) \ldots \left(\sum_{x_n} x_n \, \mathbb{P}(X_n = x_n) \right) \\
&= \mathbb{E}(X_1) \ldots \mathbb{E}(X_n) \qquad (4.123)
\end{aligned}$$

4.4.2 Covariância

Dadas duas v.a.'s X, Y no mesmo espaço de probabilidades, a *covariância* entre X e Y é definida como

$$\mathbb{C}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))], \tag{4.124}$$

isto é, é o valor esperado do produto das variáveis centradas. Expandindo o produto e usando a linearidade da esperança, obtemos

$$\begin{aligned}
\mathbb{C}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}(X\mathbb{E}(Y)) - \mathbb{E}(Y\mathbb{E}(X)) + \mathbb{E}(\mathbb{E}(Y)\mathbb{E}(X)) \\
&= \mathbb{E}[XY] - \mathbb{E}(Y)\mathbb{E}(X) - \mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y)\mathbb{E}(X) \qquad (4.125)
\end{aligned}$$

e concluimos que

$$\mathbb{C}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \tag{4.126}$$

(veja (2.113)).

No Exemplo 4.37, de (4.105-4.108), temos

$$\begin{aligned}
\mathbb{E}(XY) &= 0\times 0\,\mathbb{P}(X=0,Y=0)+0\times 1\,\mathbb{P}(X=0,Y=1)\\
&+ 1\times 0\,\mathbb{P}(X=1,Y=0)+1\times 1\,\mathbb{P}(X=1,Y=1)\\
&= \mathbb{P}(X=1,Y=1)=0.34, \qquad (4.127)
\end{aligned}$$

e

$$\begin{aligned}
\mathbb{E}(X) &= 0\,\mathbb{P}(X=0)+1\,\mathbb{P}(X=1)=\mathbb{P}(X=1)=\mathbb{P}(A_1)\overset{(3.62)}{=}0.6\\
\mathbb{E}(Y) &= 0\,\mathbb{P}(Y=0)+1\,\mathbb{P}(Y=1)=\mathbb{P}(Y=1)=\mathbb{P}(A_2)\overset{(3.66)}{=}0.6.
\end{aligned}$$

De (3.62) vem

$$\mathbb{C}(X,Y)=0.34-(0.6)^2=-0.02. \qquad (4.128)$$

Proposição 4.42 *Se X e Y forem independentes, temos*

$$\mathbb{C}(X,Y)=0. \qquad (4.129)$$

Demonstração Segue imediatamente da Proposição 4.41 e (3.62).

No Exemplo 4.39, vimos que X e Y são independentes. Logo,

$$\mathbb{C}(X,Y)=0.$$

Observação 4.43 *Dadas duas v.a.'s X e Y, uma forma de calcular $\mathbb{E}(XY)$ é a seguinte.*

$$\mathbb{E}(XY)=\mathbb{E}[Y\,\mathbb{E}(X|Y)]. \qquad (4.130)$$

Demonstração *Vamos considerar o caso discreto apenas. De (4.109),*

$$\begin{aligned}
\mathbb{E}[Y\,\mathbb{E}(X|Y)] &= \sum_{y\in\mathcal{V}_Y} y\,\mathbb{E}(X|Y=y)\,\mathbb{P}(Y=y)\\
&= \sum_{y\in\mathcal{V}_Y} y \sum_{x\in\mathcal{V}_X} x\,\mathbb{P}(X=x|Y=y)\,\mathbb{P}(Y=y)\\
&= \sum_{x\in\mathcal{V}_X}\sum_{y\in\mathcal{V}_Y} x\,y\,\mathbb{P}(X=x,\,Y=y)\\
&= \mathbb{E}(XY), \qquad (4.131)
\end{aligned}$$

onde usamos a regra do produto na penúltima passagem.

4.4.3 Soma de variáveis aleatórias

Sejam $X_1, X_2 \ldots$ v.a.'s num mesmo espaço de probabilidades e, para $n \geq 1$, seja

$$S_n = X_1 + \ldots + X_n = \sum_{i=1}^{n} X_i. \tag{4.132}$$

Somas de v.a.'s como esta entram, por exemplo, em amostragem, na estimação de médias populacionais: seja Π uma população e X uma variável numérica aí definida; seja $\mu = M(X)$. Para estimar μ, colhemos uma amostra casual simples de tamanho n de Π, e medimos X em cada indivíduo amostrado, obtendo desta forma as v.a.'s $X_1, \ldots, X_n$ (que chamamos neste caso de *amostra casual simples (de tamanho n) de X*). Então

$$\bar{X}_n := \frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i, \tag{4.133}$$

a *média amostral*, seria um *estimador* para μ.

Da linearidade da esperança, temos que

$$\mathbb{E}(S_n) = \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n) = \sum_{i=1}^{n} \mathbb{E}(X_i). \tag{4.134}$$

Vimos acima (na Observação 4.40) que, na amostragem casual simples de tamanho n com reposição, as v.a.'s $X_1, \ldots, X_n$ resultantes da medição de X nos indivíduos da amostra têm todas distribuição marginal de probabilidades dadas pela distribuição de freqüências de X. Isto também é válido (mas não tão óbvio) para a amostragem casual simples de tamanho n *sem* reposição (mas neste caso, perde-se a independência entre as v.a.'s). Logo, se $X_1, \ldots, X_n$ for uma amostra casual simples de X, temos da linearidade da esperança, de (4.134) e do que acabamos de dizer que

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{S_n}{n}\right) = \frac{1}{n}\mathbb{E}(S_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(X_i) = \mu, \tag{4.135}$$

onde a última igualdade se deve ao fato que $\mathbb{E}(X_i) = M(X) = \mu$ para todo i. Podemos então afirmar que a média amostral é um estimador *não-viesado* para μ.

Vamos em seguida considerar a variância de S_n.

Proposição 4.44 *Sejam $X_1, X_2 \ldots$ v.a.'s num mesmo espaço de probabilidades e, para $n \geq 1$, $S_n = X_1 + \ldots + X_n$. Então*

$$\mathbb{V}(S_n) = \sum_{i=1}^{n} \mathbb{V}(X_i) + 2 \sum_{\substack{i,j=1 \\ i<j}}^{n} \mathbb{C}(X_i, X_j). \tag{4.136}$$

Corolário 4.45 *No contexto da Proposição 4.44, se $X_1, X_2 \ldots$ forem independentes (duas a duas), então*

$$\mathbb{V}(S_n) = \sum_{i=1}^{n} \mathbb{V}(X_i). \tag{4.137}$$

Demonstração da Proposição 4.44

$$(S_n)^2 = \left(\sum_{i=1}^{n} X_i \right)^2 = \sum_{i=1}^{n} X_i^2 + 2 \sum_{\substack{i,j=1 \\ i<j}}^{n} X_i \, X_j, \tag{4.138}$$

logo

$$\mathbb{E}\left[(S_n)^2\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i^2\right] + \mathbb{E}\left[2 \sum_{\substack{i,j=1 \\ i<j}}^{n} X_i \, X_j\right] = \sum_{i=1}^{n} \mathbb{E}\left(X_i^2\right) + 2 \sum_{\substack{i,j=1 \\ i<j}}^{n} \mathbb{E}(X_i \, X_j). \tag{4.139}$$

Temos ainda que

$$(\mathbb{E}(S_n))^2 = \left(\sum_{i=1}^{n} \mathbb{E}(X_i) \right)^2 = \sum_{i=1}^{n} [\mathbb{E}(X_i)]^2 + 2 \sum_{\substack{i,j=1 \\ i<j}}^{n} \mathbb{E}(X_i) \, \mathbb{E}(X_j). \tag{4.140}$$

De (4.138), (4.139) e (4.24)

$$\begin{aligned} \mathbb{V}(S_n) &= \sum_{i=1}^{n} \left\{ \mathbb{E}\left(X_i^2\right) - [\mathbb{E}(X_i)]^2 \right\} + 2 \sum_{\substack{i,j=1 \\ i<j}}^{n} \left\{ \mathbb{E}(X_i \, X_j) - \mathbb{E}(X_i) \, \mathbb{E}(X_j) \right\} \\ &= \sum_{i=1}^{n} \mathbb{V}(X_i) + 2 \sum_{\substack{i,j=1 \\ i<j}}^{n} \mathbb{C}(X_i, X_j). \end{aligned} \tag{4.141}$$

Demonstração do Corolário 4.45 Imediata das Proposições 4.44 e 4.42.

Aplicações a distribuições de v.a.'s já vistas

Demonstração da Proposição 4.19

Se $X \sim \text{Bin}(n,p)$, então X pode ser escrita como

$$S_n = \sum_{i=1}^{n} Y_i, \tag{4.142}$$

onde $Y_1, \ldots, Y_n$ são v.a.'s de Bernoulli com parâmetro p independentes. De fato, se tomarmos

$$Y_i = \begin{cases} 1, & \text{se o } i\text{-ésimo ensaio de Bernoulli resulta em sucesso,} \\ 0, & \text{se o } i\text{-ésimo ensaio de Bernoulli resulta em fracasso,} \end{cases} \tag{4.143}$$

então temos claramente a validade de (4.142). A independência alegada entre as v.a.'s $Y_1, \ldots, Y_n$ segue da independência entre os ensaios.

De (4.134) e (4.36),

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(Y_i) = \sum_{i=1}^{n} p = np, \tag{4.144}$$

e (4.41) está verificada. De (4.137) e (4.37),

$$\mathbb{V}(X) = \sum_{i=1}^{n} \mathbb{V}(Y_i) = \sum_{i=1}^{n} p(1-p) = np(1-p), \tag{4.145}$$

e (4.42) está verificada.

Demonstração da Proposição 4.25 Como no caso da binomial com as Bernoulli's, uma v.a. com distribuição binomial negativa pode ser escrita como uma soma de v.a.'s geométricas de parâmetro p independentes. Isto é, se $X \sim \text{BN}(n,p)$, então

$$S_n = \sum_{i=1}^{n} Y_i, \tag{4.146}$$

onde $Y_i \sim \text{G}(p)$ para todo $i = 1, \ldots, n$, e $Y_1, \ldots, Y_n$ são independentes. Então

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(Y_i) = \sum_{i=1}^{n} \frac{1-p}{p} = n\,\frac{1-p}{p}, \tag{4.147}$$

$$\mathbb{V}(X) = \sum_{i=1}^{n} \mathbb{V}(Y_i) = \sum_{i=1}^{n} \frac{1-p}{p^2} = n\,\frac{1-p}{p^2}. \tag{4.148}$$

Aplicações na avaliação da média amostral como estimador da média populacional

Seja X uma variável populacional numérica com média $M(X) = \mu$ e variância $V(X) = \sigma^2$, e seja $X_1, \ldots, X_n$ uma amostra casual simples *com reposição* de X. Em (4.135), achamos a esperança da média amostral $\bar{X}_n$, e concluímos que se trata de estimador não-viesado para μ. De (4.25) e (4.137)

$$\mathbb{V}(\bar{X}_n) = \mathbb{V}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\,\mathbb{V}(S_n) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \qquad (4.149)$$

Note que a variância de $\bar{X}_n$ vai para 0 quando $n \to \infty$. Como a variância de $\bar{X}_n$ pode ser vista como o desvio quadrático médio da média amostral em relação à média populacional μ, concluímos que esta *distância* entre o estimador e o parâmetro estimado decai indefinidamente conforme aumentamos o tamanho da amostra. Por esta propriedade, dizemos que a média amostral (neste contexto) é um estimador *consistente* para a média populacional.

Observação 4.46 *A discussão que acabamos de fazer produz também a interpretação estatística da esperança que mencionamos na Observação 4.14. Note que média populacional μ é também a esperança comum de $X_1, \ldots, X_n$. Logo, do fato que a distância entre μ e a média de $X_1, \ldots, X_n$ vai a zero quando $n \to \infty$* [2]*, podemos dizer que a esperança de uma v.a. é o valor que esperamos observar, não necessariamente numa realização do experimento aleatório subjacente, mas como média das observações da variável em várias repetições independentes do experimento.*

Teorema Central do Limite

O resultado de aproximação normal para a distribuição binomial que vimos na Subseção 4.3.3, mais especificamente o Teorema Central do Limite (apresentado na Observação 4.34) são versões do seguinte resultado mais geral.

Teorema 4.47 (Teorema Central do Limite)
Sejam $X_1, X_2 \ldots$ v.a.'s independentes e de mesma distribuição marginal, com

[2] Distância dada pelo desvio quadrático médio, como vimos acima, mas o mesmo vale para outras distâncias.

média comum μ e variância comum $\sigma^2 > 0$. Para $n \geq 1$, seja $S_n = \sum_{i=1}^n X_i$ e $\bar{X}_n = S_n/n$. Seja

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\mathbb{DP}(S_n)} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \tag{4.150}$$

$$= \frac{\bar{X}_n - \mathbb{E}(\bar{X}_n)}{\mathbb{DP}(\bar{X}_n)} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}. \tag{4.151}$$

Então, vale a conclusão da Observação 4.34, qual seja, para todo $a, b \in \mathbb{R}$ com $a < b$, temos que

$$\lim_{n\to\infty} \mathbb{P}(a < Z_n < b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx. \tag{4.152}$$

Note que no caso da Observação 4.34, a distribuição comum de $X_1, X_2 \ldots$ é Bernoulli com parâmetro p, como observamos em (4.142) e (4.143).

Exemplo 4.48 *Suponha que a população de certa região em certa época tenha altura média 170 com desvio-padrão 10. Depois de certo tempo, desconfia-se que a altura média mudou, mas não o desvio padrão. Para estimar a altura média atual, planeja-se colher uma amostra casual simples de tamanho 100 com reposição desta população. Qual a probabilidade de a média amostral não diferir da média populacional por mais do que 2?*

Seja X a altura atual da população, e seja $\mu = M(X)$, a altura média atual. Sejam $X_1, \ldots, X_{100}$ as alturas amostradas, e $\bar{X}$ a média amostral. Então queremos achar

$$\mathbb{P}(|\bar{X} - \mu| \leq 2). \tag{4.153}$$

Temos que o desvio-padrão comum das v.a.'s da amostra é 10. Como

$$|\bar{X} - \mu| \leq 2 \ \ sse \ \ |Z| \leq \frac{2}{\sigma/\sqrt{100}} = 2,$$

onde

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{100}},$$

temos de (4.152) que

$$\mathbb{P}(|\bar{X} - \mu| \leq 2) = \mathbb{P}(|Z| \leq 2) = \mathbb{P}(-2 \leq Z \leq 2) \approx \mathbb{P}(-2 \leq \tilde{Z} \leq 2),$$

onde $\tilde{Z} \sim$N(0,1). Logo a probabilidade em questão vale aproximadamente

$$A(2) - A(-2) = 2A(2) - 1 = 2 \times 0.9773 - 1 = 0.955.$$

Observação 4.49 *A probabilidade em (4.153) neste contexto é chamada de* coeficiente de confiança *para* $\bar{X}$ *como estimador de* μ *com* margem de erro *de 2. Podemos dizer então neste caso que temos um coeficiente de confiança de aproximadamente* 95% *para* $\bar{X}$ *como estimador de* μ *com uma margem de erro de 2.*