Métodos Probabilísticos e Estatísticos com Aplicações em Engenharias e Ciências Exatas

Marcilia Andrade Campos **Leandro Chaves Rêgo**
André Feitoza de Mendonça

19 de Novembro de 2012

Prefácio

Este livro contempla um programa introdutório de Probabilidade, Estatística, Processos Estocásticos e Estatìstica Descritiva. O livro é organizado de forma tal que pode ser utilizado tanto em um curso de graduação (semestral) nas áreas de exatas e tecnologia, bem como um curso de mestrado profissionalizante destas áreas. A seção *Aprendendo um pouco mais* apresenta os tópicos mais técnicos para um curso de graduação, e são mais adequados para um estudante de pós-graduação ou para os estudantes de graduação mais curiosos.

Partes foram geradas a partir da experiência do segundo autor ministrando disciplinas de probabilidade para graduação em diversos cursos de Engenharia e no Bacharelado em Estatística da UFPE desde 2006 e também no mestrado em Estatística da UFPE desde 2007. Outras partes deste livro foram geradas a partir de cursos com este programa ministrados pela primeira autora nas graduações de Ciência da Computação e Engenharia da Computação no Centro de Informática/UFPE desde 1998. Aplicações de probabilidade na modelagem de falhas de sistemas foram introduzidas pelo terceiro autor, bem como as ilustrações dessa obra.

Como o objetivo é que possa ser usado como um livro de texto, o livro contém exercícios resolvidos e outros propostos. Respostas da alguns dos problemas propostos encontram-se ao final de cada capítulo. Muitos dos exercícios são retirados dos excelentes livros constantes nas referências.

Recife, novembro de 2012
Marcilia Andrade Campos, Leandro Chaves Rêgo & Andre Feitoza de Mendonça

Lista de Símbolos

$\mathbb{N}$	conjunto dos números naturais
Z	conjunto dos números inteiros
Z^+	conjunto dos números inteiros positivos
$\mathbb{Q}$	conjunto dos números racionais
$\mathbb{R}$	conjunto dos números reais
$\mathbb{C}$	conjunto dos números complexos
$\emptyset$	conjunto vazio
$\mathcal{P}(A)$, 2^A	conjunto das partes de A
a, b, x, y	números reais
$\vec{x}$	vetor real
$\mathcal{F}$	álgebra
$\mathcal{A}$	σ-álgebra
$\mathcal{B}$	σ-álgebra de Borel
Ω	espaço de resultados elementares, espaço amostral
ω	evento simples, resultado elementar
A, B	eventos aleatórios, eventos
A^c ou $\overline{A}$	evento complementar de A
$P(A)$	probabilidade de A
$P(A \mid B)$	probabilidade condicional de A dado B
X, Y, Z	variáveis aleatórias
$(X_1, \cdots, X_n)$ ou $X_1, \cdots, X_n$	amostra aleatória simples
iid	variáveis aleatórias independentes e identicamente distribuídas
f	função densidade
f_X	função densidade da variável aleatória X
F	função de distribuição acumulada ou função de distribuição
F_X	função de distribuição da variável aleatória X
$F_{\vec{X}}$	função de distribuição do vetor aleatório $\vec{X}$
$\vec{X}$	vetor aleatório
$\sim$	se distribui, a variável aleatória tem distribuição
$\|\|A\|\|$	cardinalidade, tamanho ou dimensão do conjunto A
∞	infinito
$\Leftrightarrow$	se e somente se
$\rightarrow$	limite ou tende
$\uparrow$	limite de seqência monotônica não-decrescente
$\downarrow$	limite de seqüência monotônoca não-crescente
$\Rightarrow$	implica

$\cap$	interseção
$\cup$	união
$\wedge$	e
$\vee$	ou
$\neg$	não
$\in$	pertence
$\notin$	não pertence
$<$	menor
$>$	maior
$\leq$	menor ou igual
$\geq$	maior ou igual
$\subseteq$	inclusão
$\subset$	inclusão estrita
$\approx$	aproximadamente igual
$\neq$	diferente
$\equiv$	equivalente
$\forall$	para todo ou qualquer que seja
$\exists$	existe
$:$	tal que
$\vert\ \vert$	valor absoluto
A_n^k, $(n)_k$	arranjo de n elementos tomados k deles
C_n^k ou $\binom{n}{k}$	combinação de n elementos tomados k deles
$!$	fatorial

Conteúdo

Capítulo 1

Probabilidade: definições, propriedades

Que ferramentas usar, e como, analisar, entender, modelar as seguintes situações:

(i) análise de tempo de execução de um algoritmo:

- pior caso (*worst-case*);
- caso médio (*average-case*).

(ii) alocamento dinâmico de memória;

(iii) análise do erro de arredondamento acumulado em um algoritmo numérico;

(iv) análise de um sistema computacional servindo a um grande número de usuários.

Modelagem dessas situações são diferentes da modelagem, por exemplo, de enviar uma aeronave ao espaço e monitorar sua órbita! Lançar uma moeda e observar a face voltada para cima, lançar um dado e observar o número na face voltada para cima, tirar uma carta do baralho e ver qual o naipe, contar o número de usuários acessando o servidor de e-mail em um dado instante, contar o número de arquivos na fila de uma impressora, entre outros, são situações cujo resultado foge ao controle do observador. Neste caso a modelagem matemática proposta é denominada de *probabilística* ou *estocástica*, a qual é realizada fundamentalmente pela Teoria da Probabilidade ou Processos Estocásticos.

Este capítulo inicia o estudo de probabilidade deste livro. Para tanto, inicialmente, antes de entrar no estudo de probabilidade, conceitos básicos sobre conjuntos são apresentados.

1.1 Conjuntos

Definição 1.1.1: Um é uma coleção de elementos distintos[1] onde os elementos não são ordenados.

[1]Na Estatística é comum se falar de conjuntos incluindo o caso onde seus elementos não são distintos. Por exemplo, o conjunto dos tempos de acesso a um banco de dados, o conjunto das notas de uma dada disciplina, entre outros, pode ter valores iguais

Esta definição intuitiva de um conjunto foi dada primeiramente por Georg Cantor (1845-1918), que criou a teoria dos conjuntos em 1895. Um conjunto pode ser especificado, listando seus elementos dentro de chaves. Por exemplo,

$$A = \{0, 1, 2, 3, 5, 8, 13\},\ B = \{0, 1, 2, \ldots, 1000\}.$$

Alternativamente, um conjunto pode ser especificado por uma regra que determina seus membros, como em:

$$C = \{x : x \text{ é inteiro e positivo}\} \text{ ou } D = \{x : x \text{ é par}\}.$$

Como em um conjunto a ordem dos elementos não importa, tem-se que:

$$\{1, 2, 3\} = \{2, 3, 1\}.$$

Se um dado elemento faz parte de um conjunto, diz-se que ele pertence ao conjunto e denota-se isso com símbolo $\in$. Por exemplo, $2 \in D = \{x : x \text{ é par}\}$ ou $3 \in E = \{x :$ é primo $\}$.

Por outro lado, se um dado elemento não faz parte de um conjunto, diz-se que ele não pertence ao conjunto e denota-se isso com o símbolo $\notin$. Por exemplo, $3 \notin D = \{x : x \text{ é par}\}$ ou $4 \notin E = \{x : x \text{ é primo}\}$.

É preciso ter cuidado ao distinguir entre um elemento como 2 e o conjunto contendo somente este elemento $\{2\}$. Enquanto, tem-se $2 \in F = \{2, 3, 5\}$, $\{2\} \notin F = \{2, 3, 5\}$, pois o conjunto contendo somente o elemento 2 não pertence à F.

Exemplo 1.1.2: Seja $G = \{2, \{3\}\}$. Então, $2 \in G$ e $\{3\} \in G$, porém $3 \notin G$. ■

O tamanho de um conjunto A, $||A||$, é a quantidade de elementos que ele possui, a qual é chamada de sua . A cardinalidade pode ser *finita*, *infinita enumerável*, ou *infinita não-enumerável*. Um conjunto é finito quando existe uma função bijetiva cujo domínio é igual a este conjunto e a imagem é o conjunto dos inteiros não-negativos menores que um número finito; seus elementos podem ser contados, sendo possível exibir seu último elemento. Um conjunto infinito enumerável tem exatamente a mesma quantidade de elementos que os naturais, ou seja, existe uma função bijetiva cujo domínio é igual a este conjunto e a imagem é igual ao conjunto dos naturais. Um conjunto é enumerável se ele for finito ou infinito enumerável. Um conjunto é não-enumerável se ele não for enumerável. Por exemplo, os seguintes conjuntos são enumeráveis:

$$N_n = \{0, 1, 2, \ldots, n-1\},$$

$$\mathcal{Z} = \{x : x \text{ é um inteiro}\},$$

$$\mathcal{Z}^+ = \{x : x \text{ é um inteiro positivo}\},$$

$$\mathcal{Q} = \{x : x \text{ é racional}\}.$$

Para notar que o conjunto dos números racionais é enumerável considere a seguinte matriz de números racionais. (Lembrando que um número x é racional se pode ser escrito sob a forma $\frac{p}{q}$, onde p e q são inteiros e $q \neq 0$.)

$$\begin{array}{cccccc}
0/1 & & 0/2 & & 0/3 & \cdots \\
 & \swarrow & & \swarrow & & \\
1/1 & & 1/2 & & 1/3 & \cdots \\
 & \swarrow & & \swarrow & & \\
2/1 & & 2/2 & & 2/3 & \cdots \\
 & \swarrow & & \swarrow & & \\
3/1 & & 3/2 & & 3/3 & \cdots \\
 & \swarrow & & \swarrow & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

Esta matriz contém todos os racionais não-negativos. Utilizando o método da diagonalização, os elementos da matriz são ordenados, sem repetição, da seguinte forma:

$$0/1, 1/1, 1/2, 2/1, 1/3, 3/1, \ldots$$

Definindo-se uma correspondência f onde para cada racional não-negativo r, $f(r)$ representa a posição em que r aparece na sequência acima, tem-se que f é uma correspondência 1-1 entre os racionais não-negativos e os naturais. Por exemplo, temos que $f(1/2) = 3$, $f(3) = 6$. Pode-se definir g no conjunto de todos os racionais tal que tal que $g(r) = 2(f(r) - 1)$ se $r > 0$, e $g(r) = 2f(|r|) - 1$ se $r \leq 0$. Desse modo, $g(r)$ é um natural par se r for um racional positivo, e um natural ímpar, se r for um racional não-positivo. Portanto, $g(r)$ é uma correspondência 1-1 entre os racionais e os naturais, o que implica que os racionais formam um conjunto enumerável.

Por outro lado, os conjuntos abaixo são não-enumeráveis:

$$\mathbb{R} = \{x : x \text{ é um número real}\},$$

$$(a, b) = \{x : a < x < b\}, \text{ onde } a < b,$$

$$[a, b] = \{x : a \leq x \leq b\}, \text{ onde } a < b.$$

Em muitos problemas o interesse é estudar um conjunto definido de objetos. Por exemplo, o conjunto dos números naturais; em outros, o conjuntos dos números reais; ou ainda, por todas as peças que saem de uma linha de produção durante um período de 24h, etc. O conjunto que contém todos os elementos objeto de estudo é chamado de *conjunto universo* e é denotado por Ω. Por outro lado, o conjunto especial que não possui elementos é chamado de *conjunto vazio* e é denotado por $\emptyset$. Este conjunto tem cardinalidade 0 e portanto é finito. Por exemplo,

$$\emptyset = \{\} = \{x : x \in \mathbb{R} \text{ e } x < x\} \text{ ou } \emptyset = (a, a).$$

Dois conjuntos A e B podem ser relacionados através da relação de inclusão, denotada por $A \subseteq B$, e lida A é um subconjunto de B ou B contém A, quando todo elemento de A é também elemento de B. Quando A não for subconjunto de B, denota-se por $A \not\subset B$. Diz-se que A é um *subconjunto próprio* de B quando se tem $A \subseteq B$, $A \neq \emptyset$, e $B \not\subset A$. Se A é subconjunto de B, então B é chamado um *superconjunto* de A. Diz-se que A e B são iguais se e somente se $A \subseteq B$ e $B \subseteq A$. Se $A \subseteq B$, então também pode-se dizer que $B \supseteq A$.

A relação $\subseteq$ possui as propriedades de (i) *reflexividade* ($A \subseteq A$); (ii) *transitividade* ($A \subseteq B, B \subseteq C \Rightarrow A \subseteq C$); e *anti-simetria* ($A \subseteq B, B \subseteq A \Rightarrow A = B$). Contudo, ela não é uma relação completa, ou seja, não é verdade que, para todos os conjuntos A e B, ou $A \subseteq B$, ou $B \subseteq A$. Também é fácil verificar que $\emptyset \subseteq A$ e $A \subseteq \Omega$ para todo conjunto A.

1.1.1 Operações com Conjuntos

Conjuntos podem ser transformados através das seguintes operações:

(i) Complementação: $A^c = \{\omega \in \Omega : \omega \notin A\}$. De acordo com esta definição, para todo $\omega \in \Omega$ e todo conjunto A, não existe outra opção além de $\omega \in A$ ou $\omega \in A^c$; além disso não pode ser verdade que $\omega \in A$ e $\omega \in A^c$ simultaneamente.

(ii) União: $A \cup B = \{\omega : \omega \in A \text{ ou } \omega \in B\}$.

(iii) Intersecção: $A \cap B = \{\omega : \omega \in A \text{ e } \omega \in B\}$.

(iv) Diferença: $A - B = A \cap B^c = \{\omega : \omega \in A \text{ e } \omega \notin B\}$.

Se $A \cap B = \emptyset$, então A e B não têm qualquer elemento em comum, e diz-se então que A e B são *disjuntos*.

Exemplo 1.1.3: Seja $\Omega = \{0, 1, 2, 3, 4, 5, 6, 7\}$, $A = \{0, 1, 5\}$ e $B = \{1, 2, 3, 4\}$. Então, $A^c = \{2, 3, 4, 6, 7\}$, $A \cup B = \{0, 1, 2, 3, 4, 5\}$, $A \cap B = \{1\}$, $A - B = \{0, 5\}$. ■

Exemplo 1.1.4: Sejam A, B, C e D subconjuntos do conjunto universo Ω tal que $A \cup B = \Omega$, $C \cap D = \emptyset$, $A \subseteq C$ e $B \subseteq D$. Prove que $A = C$ e $B = D$.
Solução: Basta provar que $C \subseteq A$ e $D \subseteq B$. Seja $\omega \in C$, então como $C \cap D = \emptyset$, tem-se que $\omega \notin D$. Logo, como $B \subseteq D$, segue que $\omega \notin B$. Mas como $A \cup B = \Omega$, tem-se que $\omega \in A$. Portanto, $C \subseteq A$.

Para provar que $D \subseteq B$, seja $\omega \in D$, então como $C \cap D = \emptyset$, tem-se que $\omega \notin C$. Logo, como $A \subseteq C$, segue que $\omega \notin A$. Mas como $A \cup B = \Omega$, tem que $\omega \in B$. Portanto, $D \subseteq B$. ■

Relações e propriedades das operações entre conjuntos incluem:

(i) Idempotência: $(A^c)^c = A$.

Prova: Suponha que $\omega \in (A^c)^c$. Então, $\omega \notin A^c$, o que por sua vez implica que $\omega \in A$, ou seja, $(A^c)^c \subseteq A$. Agora suponha que $\omega \in A$, então $\omega \notin A^c$, e portanto $\omega \in (A^c)^c$, ou seja, $A \subseteq (A^c)^c$. Logo, $(A^c)^c = A$. ▮

(ii) Comutatividade (Simetria): $A \cup B = B \cup A$ e $A \cap B = B \cap A$.

Prova: Suponha que $\omega \in A \cup B$. Então, $\omega \in A$, ou $\omega \in B$, o que implica que $\omega \in B \cup A$, ou seja, $A \cup B \subseteq B \cup A$. Agora suponha que $\omega \in B \cup A$. Então, $\omega \in B$, ou $\omega \in A$, o que por sua vez implica que $\omega \in A \cup B$, ou seja, $B \cup A \subseteq A \cup B$. Portanto, $A \cup B = B \cup A$.

A prova para o caso da intersecção é análoga e deixada como Exercício. ▮

(iii) Associatividade: $A \cup (B \cup C) = (A \cup B) \cup C$ e $A \cap (B \cap C) = (A \cap B) \cap C$.

Prova: Exercício. ▮

(iv) Distributividade: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ e $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Prova: Exercício. ▮

(v) Leis de De Morgan: $(A \cup B)^c = A^c \cap B^c$ e $(A \cap B)^c = A^c \cup B^c$.

Prova: Suponha que $\omega \in (A \cup B)^c$. Então, $\omega \notin (A \cup B)$, o que por sua vez implica que $\omega \notin A$ e $\omega \notin B$. Logo, $\omega \in A^c$ e $\omega \in B^c$, ou seja, $\omega \in (A^c \cap B^c)$. Então, $(A \cup B)^c \subseteq (A^c \cap B^c)$. Agora suponha que $\omega \in (A^c \cap B^c)$. Então, $\omega \in A^c$ e $\omega \in B^c$, o que por sua vez implica que $\omega \notin A$ e $\omega \notin B$. Logo, $\omega \notin (A \cup B)$, ou seja, $\omega \in (A \cup B)^c$. Então, $(A^c \cap B^c) \subseteq (A \cup b)^c$. Portanto, $(A^c \cap B^c) = (A \cup b)^c$.

A prova da outra Lei de De Morgan é análoga e deixada como exercício. ▮

As Leis de De Morgan permitem que se possa expressar uniões em termos de intersecções e complementos e intersecções em termos de uniões e complementos.

Uniões e intersecções podem ser estendendidas para coleções arbitrárias de conjuntos. Seja $\mathcal{I}$ um conjunto qualquer. Este conjunto $\mathcal{I}$ será utilizado para indexar, ou seja, identificar através de um único símbolo os conjuntos na coleção arbitrária de interesse e desse modo simplificar a notação utilizada. Por exemplo, se $\mathcal{I} = \{1, 5, 7\}$, então $\cup_{i \in \mathcal{I}} A_i = A_1 \cup A_5 \cup A_7$; ou, se $\mathcal{I} = \mathbb{N}$, então $\cap_{i \in \mathbb{N}} A_i = A_1 \cap A_2 \cap \cdots \cap A_n \cdots$.

De modo análogo ao caso de dois conjuntos, define-se:

$$\cup_{i \in \mathcal{I}} A_i = \{\omega \in \Omega : \omega \text{ pertence a pelo menos um dos conjuntos } A_i, \text{ onde } i \in \mathcal{I}, \}$$

e

$$\cap_{i \in \mathcal{I}} A_i = \{\omega \in \Omega : \omega \text{ pertence a todo } A_i, \text{ onde } i \in \mathcal{I}.\}$$

Se $\mathcal{I}$ for um conjunto enumerável, diz-se que $\cup_{i \in \mathcal{I}} A_i$, respectivamente, $\cap_{i \in \mathcal{I}} A_i$, é uma união, respectivamente intersecção, enumerável de conjuntos.

Exemplo 1.1.5: Se $A_i = [1, 2 + \frac{1}{i})$, $i \in \mathbb{N}$, então $\cup_{i \in \mathbb{N}} A_i = [1, 3)$ e $\cap_{i \in \mathbb{N}} = [1, 2]$. ■

1.1.2 Produto Cartesiano

Definição 1.1.6: Produto Cartesiano. O produto Cartesiano $A \times B$ de dois conjuntos dados A e B é o conjunto de todos os pares ordenados de elementos, onde o primeiro pertence à A e o segundo pertence à B:

$$A \times B = \{(a, b) : a \in A, b \in B\}.$$

Por exemplo, se $A = \{a, b, c\}$ e $B = \{1, 2\}$:

$$A \times B = \{(a, 1), (a, 2), (b, 1), (b, 2), (c, 1), (c, 2)\}$$

e

$$B \times A = \{(1, a), (1, b), (1, c), (2, a), (2, b), (2, c)\}.$$

O produto cartesiano de dois conjuntos pode ser estendido para n conjuntos da seguinte maneira: se $A_1, \ldots, A_n$ forem conjuntos, então,

$$A_1 \times A_2 \times \ldots \times A_n = \{(a_1, a_2, \ldots, a_n) : a_i \in A_i\},$$

ou seja, o conjunto de todas as ênuplas ordenadas.

Um caso especial importante é o produto cartesiano de um conjunto por ele próprio, isto é, $A \times A$. Exemplos disso são o plano euclideano, $\mathbb{R} \times \mathbb{R}$, e o espaço euclideano tridimensional, representado por $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$.

1.1.3 Conjunto das Partes

Definição 1.1.7: Dado um conjunto qualquer A, pode-se definir um outro conjunto, conhecido como *conjuntos das partes de* A, e denotado por 2^A, cujos elementos são subconjuntos de A.

Exemplo 1.1.8: Seja $A = \{1, 2, 3\}$, então

$$2^A = \{\emptyset, A, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\}.$$

Pode-se provar que a cardinalidade do conjunto das partes de qualquer conjunto dado A é maior que a cardinalidade de A.

Teorema 1.1.9: *Se A é um conjunto e 2^A é o conjunto das partes de A, não existe uma função $f : A \to 2^A$ que seja sobrejetiva.*

Prova: Recorde que uma função $g : D \to I$ é sobrejetiva se para todo $y \in I$, existe $x \in D$ tal que $g(x) = y$. Suponha por contradição, que existe uma função sobrejetiva $f : A \to 2^A$.

Defina o conjunto, $B = \{x \in A : x \notin f(x)\}$. Como f por suposição é sobrejetiva e $B \in 2^A$, tem-se que existe $b \in A$ tal que $f(b) = B$. Existem dois casos a considerar: $b \in B$ ou $b \in B^c$. Se $b \in B$, então $b \notin f(b)$. Mas como $B = f(b)$, tem-se que $b \notin B$, absurdo. Se $b \in B^c$, então $b \in f(b)$. Mas como $B = f(b)$, tem-se que $b \in B$, absurdo. ∎

1.1.4 Partição

Intuitivamente, uma partição de um conjunto universo é uma maneira de distribuir os elementos deste conjunto em uma coleção arbitrária de subconjuntos. Formalmente, tem-se a seguinte definição:

Definição 1.1.10: Dado um conjunto universo Ω, uma partição $\Pi = \{A_\alpha, \alpha \in \mathcal{I}\}$ de Ω é uma coleção de subconjuntos de Ω (neste caso, indexados por α que toma valores no conjunto de índices $\mathcal{I}$) e satisfaz:

(i) Para todo $\alpha \neq \beta$, $A_\alpha \cap A_\beta = \emptyset$;

(ii) $\cup_{\alpha \in \mathcal{I}} A_\alpha = \Omega$.

Deste modo os conjuntos de uma partição são disjuntos par a par e cobrem todo o conjunto universo. Portanto, cada elemento $\omega \in \Omega$ pertence a um, e somente um, dos conjuntos A_α de uma partição.

Exemplo 1.1.11: Se $\Omega = \{1, 2, 3, 4\}$, então $\{A_1, A_2\}$, onde $A_1 = \{1, 2, 3\}$ e $A_2 = \{4\}$, é uma partição de Ω. ■

Exemplo 1.1.12: A coleção de intervalos $\{(n, n+1] : n \in Z\}$ é uma partição dos números reais $\mathbb{R}$. ■

1.1.5 Função Indicadora

É sempre conveniente representar um conjunto A por uma função I_A tendo domínio (conjunto dos argumentos da função) Ω e contra-domínio (conjunto dos possíveis valores da função) binário $\{0, 1\}$.

Definição 1.1.13: **Função Indicadora.** A função indicadora $I_A : \Omega \to \{0, 1\}$ de um conjunto A é dada por

$$I_A(\omega) = \begin{cases} 1, & \text{se } \omega \in A, \\ 0, & \text{se } \omega \notin A. \end{cases}$$

É fácil observar que $I_\Omega(\omega) = 1, \forall \omega \in \Omega$ e que $I_\emptyset(\omega) = 0, \forall \omega \in \Omega$. Note que existe uma correspondência 1-1 entre conjuntos e suas funções indicadoras:

$$A = B \Leftrightarrow (\forall \omega \in \Omega) I_A(\omega) = I_B(\omega).$$

Esta relação entre conjuntos e suas respectivas funções indicadoras permite usar a aritmética de funções indicadoras para explorar propriedades de conjuntos. É fácil provar as seguintes propriedades de funções indicadoras:

$$I_{A^c} = 1 - I_A,$$

$$A \subseteq B \Leftrightarrow I_A \leq I_B,$$

$$I_{A \cap B} = \min(I_A, I_B) = I_A I_B,$$

$$I_{A \cup B} = \max(I_A, I_B) = I_A + I_B - I_{A \cap B},$$

$$I_{A-B} = \max(I_A - I_B, 0) = I_A I_{B^c}.$$

Exemplo 1.1.14: Utilizando funções indicadoras, verifique que $A \subseteq B \Leftrightarrow B^c \subseteq A^c$.
Solução: Tem-se que

$$A \subseteq B \Leftrightarrow I_A \leq I_B \Leftrightarrow 1 - I_A \geq 1 - I_B \Leftrightarrow I_{A^c} \geq I_{B^c} \Leftrightarrow B^c \subseteq A^c.$$

■

1.2 Breve Histórico sobre o Estudo da Chance e da Incerteza

Antes de começar as definições e propriedades da função probabilidade, será dado um breve histórico a partir do século XVI.

...a era dos jogos de azar...

Cardano (1501-1576).

Primeiro matemático que calculou uma probabilidade corretamente. Introduziu a idéia de combinações para calcular o cardinal do espaço amostral e do número de eventos elementares favoráveis, de modo que o quociente entre ambos os números desse um resultado que estivesse de acordo com a experiência.

Fermat (1601-1655), Pascal (1623-1662), Huygens (1629-1695).

Um dos primeiros problemas interessantes em probabilidade foi proposto pelo nobre francês Chevalier de Méré. O problema é o seguinte: dois jogadores, A e B, concordam em jogar uma série de jogos, s. Por alguma razão acidental, eles decidem parar o jogo quando A tem ganho m jogos e B, n, sendo $m \leq s$, $n \leq s$ e $m \neq n$. A pergunta é: como as apostas devem ser divididas?

A solução desse problema envolveu Fermat, Pascal e Huygens.

Huygens (1629-1695).

Huygens publicou em 1657 o primeiro livro sobre Teoria da Probabilidade *De Ratiociniis in Alae Ludo* (On Calculations in Game of Chance), o qual foi muito bem aceito pelos matemáticos da época e foi a única introdução à Teoria da Probabilidade durante 50 anos.

Ainda datam desse período os fundamentos do conceito de esperança matemática, o teorema da adição de probabilidades e o teorema da multiplicação de probabilidades.

...o começo...

James Bernoulli (1654-1705).

Publicou em 1713 *Ars Conjectandi* (The Art of Guessing), obra dividida em quatro partes, onde, na última, provou o primeiro limite da Teoria da Probabilidade, *A Lei dos Grandes Números*, ou Teorema de Ouro.

Pierre Simon, Marquês de Laplace (1749-1827).

Publicou em 1812 *Théorie Analytique des Probabilités*, no qual apresentou seus próprios resultados e os de seus predecessores. Suas contribuições mais importantes foram a (i) aplicação de métodos probabilísticos aos erros de observações e (ii) formulou a idéia de considerar os erros de observações como o resultado acumulativo da adição de um grande número de erros elementares independentes.

Poisson (1781-1840), Gauss (1777-1855).

Ambos tiveram grande interesse por teoremas limite. A Gauss é creditada a origem da Teoria dos Erros, em particular, dos Mínimos Quadrados.

...estamos chegando...

P. L. Chebyshev (1822-1894), A. A. Markov (1856-1922), A. M. Lyapunov (1857-1918).

Desenvolveram métodos efetivos para provar teoremas limite para soma de variáveis aleatórias independentes, mas arbitrariamente distribuídas. Chebyshev foi o primeiro a explorar com profundidade as relações entre variáveis aleatórias e suas esperanças matemáticas.

As contribuições de Markov, relacionam-se com teoremas limite para soma de variáveis aleatórias independentes e a criação de um novo ramo da Teoria da Probabilidade: a teoria das variáveis aleatórias dependentes conhecidas como Cadeias de Markov.

Uma das contribuições de Lyapunov foi o uso da função característica para provar o teorema central do limite.

John von Neumann (1903-1957).

Ainda nessa época, von Neumann assentou sobre bases firmes a *Teoria dos Jogos*, em 1928, contribuiu para a descoberta da Mecânica Quântica, contribuiu para o desenvolvimento da primeira bomba atômica americana e **...inventou o computador digital!**

...a axiomatização...

Lebesgue, definiu a Teoria da Medida e Integração.

Borel (1871-1956), estabeleu a analogia entre medida de um conjunto e probabilidade de um evento e integral de uma função e esperança matemática.

A. N. KOLMOGOROV, publicou em 1933 *Foundations of the Theory of Probability*, com a axiomática que tem sido usada até hoje.

...hoje...

Atualmente, idéias recentes em Teoria da Probabilidade são (i) Probabilidade Intervalar (Campos, 1997)(Santos, 2010), (ii) Probabilidades Imprecisas[2] e (iii) Probabilidade sobre Domínios[3].

1.3 Experimento Aleatório

Um *experimento* é qualquer processo de observação. Em muitos experimentos de interesse, existe um elemento de incerteza, ou chance, que não importa o quanto se saiba sobre o passado de outras performances deste experimento, não é possível predizer o seu comportamento em futuras realizações por várias razões: impossibilidade de saber todas as causas envolvidas; dados insuficientes sobre as suas condições iniciais; os fenômenos que o geraram podem ser tão complexos que impossibilitam o cálculo do seu efeito combinado; ou, na verdade, existe alguma aleatoriedade fundamental no experimento. Tais experimentos são conhecidos como *experimentos aleatórios*. Salvo mencionado em contrário, este livro restringe-se à classe de experimentos aleatórios cujo conjuntos de possíveis resultados seja conhecido[4].

Os resultados de um experimento aleatório são caracterizados pelos seguintes componentes:

(i) um conjunto de resultados que podem ocorrer: Ω;

(ii) uma coleção que contém como elementos conjuntos de resultados que se pretende analisar: $\mathcal{A}$;

(iii) para cada conjunto da coleção $\mathcal{A}$ um valor numérico, p, que descreve quão verossímil é a ocorrência de um resultado contido em $\mathcal{A}$.

1.4 Espaço Amostral

O conjunto de possíveis resultados de um experimento aleatório é chamado de *espaço amostral*. Em um dado experimento aleatório a especificação do espaço amostral deve ser tal que este (i) liste todos os possíveis resultados do experimento sem duplicação e o (ii) faça em um nível de detalhamento suficiente para os interesses desejados, omitindo resultados que, embora logicamente ou fisicamente possíveis, não tenham qualquer implicação prática na sua análise.

Por exemplo, suponha que o prêmio de uma determinada aposta envolvendo um jogo de baralho dependa somente da cor do naipe da carta retirada. Então, o espaço amostral poderia

[2] `http://www.sipta.org/documentation`.

[3] `http://www.it.iitb.ac.in/papers/pkdd07.pdf`, `http://dauns.math.tulane.edu/ mwm/ftp/darms.pdf`.

[4] É importante ressaltar que frequentemente são encontradas situações práticas onde não se consegue descrever todos os possíveis resultados de um experimento. Uma maneira de contornar este problema é assumir que um resultado possível do experimento é a não ocorrência de qualquer dos resultados descritos, contudo, em problemas práticos, tal suposição pode acarretar em dificuldades quando se tenta elicitar ou eduzir probabilidades de especialistas.

ser dado por $\Omega = \{\text{vermelho, preto}\}$ ou $\Omega = \{\text{copas, ouro, espada, paus}\}$ ou ainda $\Omega = \{\text{todas as 52 cartas do baralho}\}$. Portanto, muito mais se poderia dizer sobre o resultado de uma retirada de carta de um baralho que os simples resultados binários vermelho ou preto. Informações outras são ignoradas quando se usa a hipótese adicional que existe uma aposta com pagamentos que dependem apenas da cor do naipe da carta retirada.

1.5 Eventos e Coleção de Eventos

Um *evento* é um subconjunto do espaço amostral, ou seja, é um conjunto de resultados possíveis do experimento aleatório. Ao se realizar um experimento aleatório, se o resultado pertence a um dado evento A, diz-se que A *ocorreu*.

Definição 1.5.1: Os eventos A e B são *disjuntos* ou *mutuamente excludentes* ou *mutuamente exclusivos* se não puderem ocorrer juntos, ou, em liguagem de conjuntos, $A \cap B = \emptyset$.

A ocorrência de eventos combinados também é um evento; essas combinações podem ser expressas através das operações de conjuntos: complementar, união, intersecção e diferença.

Exemplo 1.5.2: Sejam A, B, e C eventos em um mesmo espaço amostral Ω. Expresse os seguintes eventos em função de A, B, e C e operações Booleanas de conjuntos.

(a) Pelo menos um deles ocorre:
$$A \cup B \cup C.$$

(b) Exatamente um deles ocorre:
$$(A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C).$$

(c) Apenas A ocorre:
$$(A \cap B^c \cap C^c).$$

(d) Pelo menos dois ocorrem:
$$(A \cap B \cap C^c) \cup (A \cap B^c \cap C) \cup (A^c \cap B \cap C) \cup (A \cap B \cap C).$$

(e) No máximo dois deles ocorrem:
$$(A \cap B \cap C)^c.$$

(f) Nenhum deles ocorre:
$$(A^c \cap B^c \cap C^c).$$

(g) Ambos A e B ocorrem, mas C não ocorre:
$$(A \cap B \cap C^c).$$

■

De acordo com Fine (2006), embora possa-se pensar que, dado um espaço amostral, necessariamente é de interesse analisar todos os seus subconjuntos (e isto eventualmente é verdadeiro), tem-se três razões para esperar que o interesse seja apenas por alguns de seus subconjuntos. Primeiro, o espaço amostral pode conter um grau de detalhamento superior ao de interesse no problema. Por exemplo, ele pode representar uma única jogada de um dado mas o interesse é apenas em saber se o resultado é par ou ímpar. Segundo, o objetivo é associar a cada evento A uma probabilidade $P(A)$; como essas probabilidades estão baseadas em algum conhecimento sobre a tendência de ocorrer o evento, ou no grau de crença que determinado evento ocorrerá, o conhecimento sobre P pode não se estender para todos os subconjuntos de Ω. A terceira (e técnica) razão para limitar a coleção de eventos de interesse é que condições impostas em P pelos axiomas de Kolmogorov, que serão vistos adiante, podem não permitir que P seja definida em todos os subconjuntos de Ω, em particular isto pode ocorrer quando Ω for não enumerável (fato este fora do escopo deste livro (Burrill, 1972)(Shiryayev, 1984).

Em probabilidade, o interesse é em uma coleção especial $\mathcal{A}$ de subconjuntos do espaço amostral Ω ($\mathcal{A}$ é um conjunto cujos elementos também são conjuntos!) que são eventos de interesse no que se refere ao experimento aleatório $\mathcal{E}$ e os quais tem-se conhecimento sobre a sua probabilidade. $\mathcal{A}$ é chamado de uma *σ-álgebra de eventos*. O domínio de uma medida de probabilidade é uma σ-álgebra.

Definição 1.5.3: Uma álgebra de eventos $\mathcal{F}$ é uma coleção de subconjuntos do espaço amostral Ω que satisfaz:

(i) $\mathcal{F}$ é não vazia;

(ii) $\mathcal{F}$ é fechada com respeito a complementos (se $A \in \mathcal{F}$, então $A^c \in \mathcal{F}$);

(iii) $\mathcal{F}$ é fechada com respeito a uniões finitas (se $A, B \in \mathcal{F}$, então $A \cup B \in \mathcal{F}$).

Definição 1.5.4: Uma *σ-álgebra* $\mathcal{A}$ é uma álgebra de eventos que também é fechada com relação a uma união enumerável de eventos,

$$(\forall i \in \mathcal{I}) A_i \in \mathcal{A} \Rightarrow \cup_{i \in \mathcal{I}} A_i \in \mathcal{A}.$$

Pelas Leis de De Morgan, tem-se que $\mathcal{A}$ também é fechada com respeito a intersecções enumeráveis.

Dado um espaço amostral Ω, a maior σ-álgebra é o conjunto das partes de Ω, $\mathcal{A} = \mathcal{P}(\Omega)$, e a menor é $\mathcal{A} = \{\emptyset, \Omega\}$. [5]

1.6 Fundamentos de Probabilidade

Raciocínio probabilístico é frequente no dia a dia da humanidade e pode ser observado tanto através de ações (atravessar uma rua, embarcar num avião ou navio, ligar um computador são atividades que revelam um julgamento de que a probabilidade de um acidente é

[5] Em geral, dada qualquer partição enumerável de Ω, pode-se definir uma nova σ-álgebra de Ω considerando todas as possíveis uniões enumeráveis de conjunto que formam a partição.

suficientemente pequena para assegurar a vantagem de se tomar tais ações) ou de palavras ("provavelemente choverá amanhã", "existe uma probabilidade suficientemente grande de que o Brasil seja o próximo campeão mundial de futebol").

De acordo com Fine (2006), o raciocínio probabilístico pode ser classificado nas seguintes dimensões:

- grau de precisão – o conceito estrutural;
- o significado, ou interpretação a ser dada à probabilidade;
- estrutura matemática formal da função probabilidade dada por um conjunto de axiomas.

O conceito estrutural determina a precisão esperada de que probabilidade represente fenômenos aleatórios. A interpretação proporciona a base com a qual a probabilidade deve ser determinada e indica o que se pode aprender com ela, ou seja, o que uma afirmação probabilística significa. O conceito estrutural e a interpretação guiam a escolha dos axiomas. O conjunto de axiomas, contudo, pode somente capturar uma parte do que se entende da interpretação.

A compreensão de fundamentos de probabilidade é importante, pois aplicações de teoria da probabilidade dependem fortemente de seus fundamentos. Por exemplo, os fundamentos influem na escolha dos métodos estatísticos a serem utilizados (frequentistas e Bayesianos, entre outros) e na interpretação dos resultados obtidos. Os próximos exemplos motivam a importância do estudo de fundamentos de probabilidade.

Exemplo 1.6.1: (Halpern, (2003)) Suponha que Alice tenha uma moeda honesta e que ela e João saibam que a moeda é honesta. Alice joga a moeda e olha o resultado. Após a moeda ser jogada, qual a probabilidade de cara segundo João? Um argumento diria que a probabilidade ainda é $1/2$, pois João nada aprendeu sobre o resultado da jogada, então ele não deve alterar o valor de sua probabilidade. Um outro argumento, questiona se realmente faz sentido falar sobre probabilidade de cara depois que a moeda foi jogada. Segundo este argumento, a moeda ou caiu cara ou coroa, então o melhor que João pode afirmar é que a probabilidade de cara ou é 0 ou é 1, mas ele não sabe discernir entre esses valores. ■

Exemplo 1.6.2: (Halpern, (2003)) Suponha agora que Alice tenha duas moedas, uma honesta e outra tendenciosa e é duas vezes mais provável dar cara que coroa com esta moeda. Alice escolhe uma das moedas (suponha que ela sabe distinguir as moedas) e está prestes a jogá-la. João sabe que uma moeda é honesta e que a outra é tendenciosa e que é duas vezes mais provável cair cara que coroa com a moeda tendenciosa, mas ele não sabe qual moeda Alice escolheu nem lhe foi dada a probabilidade com que Alice escolhe a moeda honesta. Qual a probabilidade de cara segundo João? ■

Exemplo 1.6.3: *Paradoxo de Ellsberg* (1961). Suponha que existam duas urnas cada uma com 60 bolas. A urna 1 contém 30 bolas azuis e 30 bolas verdes. Tudo que se sabe sobre a urna 2 é que ela contém bolas azuis e verdes, mas não se sabe a distribuição das bolas. Considere que existem duas loteria com prêmios baseados no sorteio de bolas dessas urnas. Loteria L_1 paga R$1.000,00 se uma bola azul for sorteada na urna 1, e R$0,00 caso contrário. Loteria L_2 paga R$1.000,00 se uma bola azul for sorteada na urna 2, e R$0,00 caso contrário. A maioria das pessoas quando questionada se prefere um bilhete da Loteria L_1 ou L_2 prefere um bilhete da loteria L_1. Suponha agora que temos duas outras loterias L_3 e L_4, onde a primeira paga R$1.000,00 somente se uma bola verde for sorteada da urna 1, e a segunda para R$1.000,00 somente se uma bola verde for sorteada da urna 2. Também, é verificado que a maioria das pessoas que preferiram a loteria L_1 à loteria L_2 preferem a loteria L_3 à loteria L_4. Com estas preferências, não é possível que o decisor possua uma única distribuição de probabilidade subjetiva sobre as cores das bolas na urna 2, pois a primeira preferência (L_1 sobre L_2) indica que o decisor considera que existam mais bolas verdes que azuis na urna 2, e a segunda (L_3 sobre L_4) indica que o decisor considera que existam mais bolas azuis que verdes na urna 2. Esse fenômeno é conhecido na literatura como *aversão a ambiguidade*, e pode-se modelar a incerteza do decisor por um conjunto de medidas de probabilidade ao invés de uma única medida de probabilidade. ■

1.6.1 Hierarquia de Conceitos Estruturais de Probabilidade

A seguir apresenta-se uma variedade de conceitos estruturais e interpretações de probabilidade que foram descritos em Fine (2005).

Possivelmente. "Possivelmente A" é o conceito mais rudimentar e menos preciso, e o usado pelos antigos Gregos para distinguir entre o que era necessário e o que era contingente. Existe um número de conceitos de possibilidade que incluem os seguintes:

- **possibilidade lógica,** no sentido que não se contradiz logicamente; ou seja, não é possível logicamente que algo ocorra e não ocorra simultaneamente.
- **possibilidade epistêmica,** segundo a qual a ocorrência de A não contradiz o conhecimento, que inclui, mas estende mais que mera lógica; por exemplo, não é epistemicamente possível que ocorra uma tempestade de neve em Recife.
- **possibilidade física,** a ocorrência de A é compatível com leis físicas, contudo pode ser extremamente improvável — por exemplo, uma moeda parando e ficando equilibrada na borda em uma superfície rígida;
- **possibilidade prática,** a noção do dia-a-dia segundo a qual A é praticamente possível se ele tem pelo menos uma verossimilhança não tão pequena de ocorrer.

Provavelmente. Provavelmente A é um fortalecimento da noção de possibilidade significando mais que provável que não provável. Enquanto ela pode corresponder ao caso que a probabilidade numérica de A seja maior que $1/2$, este conceito não requer qualquer comprometimento com uma probabilidade numérica nem com o preciso estado de conhecimento que uma probabilidade numérica requer.

Probabilidade Comparativa. "A é pelo menos tão provável quanto B". A probabilidade comparativa inclui "provavelmente A" através de "A é pelo menos tão provável quanto

A^c". Pode ser relacionada com probabilidade numérica através de $P(A) \geq P(B)$; embora como nos dois exemplos anteriores, probabilidade comparativa não requer qualquer comprometimento com probabilidade numérica.

Probabilidade Intervalar. "A tem probabilidade intervalar, ou probabilidade inferior e superior $(\underline{P}(A), \overline{P}(A))$". Isto permite um grau de indeterminação variável sem o comprometimento de que exista um "verdadeiro" valor no intervalo; além dessa probabilidade intervalar, existe outra (Campos, 1997), baseada na matemática intervalar (Moore, 1966 e 1979) e na aritmética de exatidão máxima (Kulisch & Miranker, 1981), a qual consiste de um intervalo fechado de números reais, $[\underline{P}(A), \overline{P}(A)]$ com a precisão (Sterbenz, 1974) tão pequena quanto possível.

Probabilidade Numérica. "A probabilidade de A é o número real $P(A)$." Este é o conceito usual e será o enfocado neste livro. Enquanto este conceito absorveu quase toda a atenção de pessoas envolvidas com fenômenos de chance e incerteza e provou ser frutífero na prática científica, este não é o único conceito utilizado em linguagem ordinária e no raciocínio probabilístico do dia-a-dia.

De agora em diante o foco é o conceito estrutural mais utilizado que é a probabilidade numérica.

1.6.2 Interpretações de Probabilidade

Ainda segundo Fine (2006), parece não ser possível reduzir probabilidade a outros conceitos; ela é uma noção em si mesma. O que pode ser feito é relacionar probabilidade a outros conceitos através de uma interpretação. Os cinco mais comuns grupos de interpretação para probabilidade são os seguintes:

1. Lógica: grau de confirmação da hipótese de uma proposição que "A ocorre" dada uma evidência através da proposição que "B ocorreu". Quando as evidências, ou premissas, são insuficientes para deduzir logicamente a hipótese ou conclusão, pode-se ainda medir quantitativamente o grau de suporte que uma evidência dá a uma hipótese através de probabilidade lógica. Por exemplo, um jurado tem de utilizar julgamento que envolvem probabilidades lógicas para condenar ou não um determinado réu baseado nas evidências disponíveis.

2. Subjetiva: se refere ao grau de crença pessoal na ocorrência do evento A e é medida através da interpretação comportamental de disposição a apostar ou agir. Por exemplo, se um torcedor de futebol acredita que seu time tem mais de 50% de chance de ganhar o campeonato, ele deverá preferir um bilhete de loteria que lhe pague um prêmio L se seu time for campeão a um outro bilhete que lhe pague um prêmio L à obtenção de cara no lançamento de uma moeda honesta.

3. Frequentista: se refere ao limite da frequência relativa de ocorrência do evento A em repetidas realizações não relacionadas do experimento aleatório $\mathcal{E}$. Note que limites de frequência relativas são uma idealização, pois não se pode realizar infinitas realizações de um experimento.

4. **Propensidade:** tendência, propensidade, ou disposição para um evento A ocorrer. Por exemplo, considerações de simetria, podem levar a conclusão que um dado tem a mesma propensão, ou tendência, a cair em qualquer uma de suas faces.

5. **Clássica:** baseada em uma enumeração de *casos igualmente prováveis.*

Neste livro adota-se a abordagem tradicional de interpretação de probabilidade, isto é, a frequentista.

1.7 Frequência Relativa

Para discutir sobre a terceira dimensão do raciocínio probabilístico, isto é, a associação de uma medida numérica a eventos a qual representa a probabilidade com que eles ocorrem, será utilizada a interpretação frequentista para motivar os axiomas de Kolmogorov. Fixando uma sequência de resultados $(\omega_i \in \Omega)_{i\in\mathbb{N}}$, se o interesse é na ocorrência de um dado evento A, a frequência relativa de A nada mas é que uma média aritmética da função indicadora de A calculada em cada um dos termos ω_i, ou seja,

Definição 1.7.1: A frequência relativa de um evento A, $f_n(A)$, determinada pela sequência de resultados $(\omega_1, \ldots, \omega_n)$ de n experimentos aleatórios, é

$$f_n(A) = \frac{1}{n}\sum_{i=1}^{n} I_A(\omega_i) = \frac{N_n(A)}{n},$$

sendo $N_n(A)$ o número de ocorrências de A nas n repetições do experimento aleatório.

Propriedades da frequência relativa são:

(i) $f_n : \mathcal{A} \to \mathbb{R}$.

(ii) $f_n(A) \geq 0$.

(iii) $f_n(\Omega) = 1$.

(iv) Se A e B são disjuntos, então $f_n(A \cup B) = f_n(A) + f_n(B)$.

(v) Se $A_1, A_2, \cdots A_n, \cdots$ é uma seqüência de eventos disjuntos dois a dois, então $f_n(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} f_n(A_i)$.

No que se segue, supõe-se que existe alguma base empírica, física ou sobrenatural, que garanta que $f_n(A) \to P(A)$, embora que o sentido de convergência quando n cresce só será explicado pela Lei dos Grandes Números (estudada posteriormente). Esta tendência da frequência relativa de estabilizar em um certo valor é conhecida como *regularidade estatística.* Deste modo, dada a interpretação frequentista de probabilidade é natural que P satisfaça propriedades similares às (i)-(v).

1.8 Axiomas de Kolmogorov

Antes de um sistema computacional ou algoritmo ser analisado, várias distribuições de probabilidade têm de ser analisadas. De onde vêm essas distribuições? Como é possível avaliar a vazão (*throughput*), tempo de resposta (*response time)*, confiabilidade (*reliability*) e disponibilidade (*availability*) de um sistema de comunicação? Estas e outras perguntas estão ligadas a problemas de avaliação de desempenho, a qual é suportada, primordialmente, por Probabilidade, Estatística e Processos Estocásticos.

Questões de probabilidade em situações práticas basicamente constituem-se, como seria o esperado, em como calcular probabilidades. Aí é onde a situação se complica. Se o espaço amostral é finito e os resultados equiprováveis, pode-se usar a definição clássica e a "complicação" consiste em contar, o que implica no uso de técnicas de análise combinatória, que, não são fáceis. Se o problema envolve "volumes de sólidos", é possível, em algumas situações, usar as chamadas probabilidades geométricas e o problema está resolvido. Se o espaço amostral é enumerável, conhecimentos sobre progressões geométricas adquiridos no segundo grau resolvem alguns problemas. Uma outra forma para calcular probabilidades é usar a frequência relativa como sendo a probabilidade para um dado evento. Nesse caso teríamos que ter um "grande número de observações", mas, o que é $\cdots$ "grande"? Portanto a construção axiomática da teoria da probabilidade, abstrai o cálculo de probabilidades de casos particulares e nos provê de um método formal para resolver problemas probabilísticos.

Os axiomas descritos a seguir não descrevem um único modelo probabilístico, apenas determinam quais funções matemáticas podem ser usadas para se descrever as probabilidades envolvidas em um experimento qualquer. A partir dos axiomas pode-se utilizar métodos matemáticos para encontrar propriedades que serão verdadeiras em qualquer modelo probabilístico. A escolha de um modelo específico satisfazendo os axiomas é feita pelo probabilista, ou estatístico, familiar com o fenômeno aleatório sendo modelado.

Os primeiros quatro axiomas de Kolmogorov são diretamente relacionados com as propriedades de frequência relativa.

(K1) Inicial. O experimento aleatório é descrito pelo espaço de probabilidade $(\Omega, \mathcal{A}, P)$ que consiste do espaço amostral Ω, de uma σ-álgebra $\mathcal{A}$, construída a partir de Ω e de uma função de valores reais $P : \mathcal{A} \to \mathbb{R}$.

(K2) Não-negatividade. $\forall A \in \mathcal{A}, P(A) \geq 0$.

(K3) Normalização Unitária. $P(\Omega) = 1$.

(K4) Aditividade Finita. Se A, B são disjuntos, então $P(A \cup B) = P(A) + P(B)$.

É fácil provar (tente!) utilizando indução matemática que (K4) é válida para qualquer coleção finita de eventos disjuntos dois a dois, ou seja, se $A_i \cap A_j = \emptyset$, $\forall i \neq j$, com $i, j = 1 \cdots n$, então $P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.

Um quinto axioma, embora não tenha significado em espaços amostrais finitos, foi proposto por Kolmogorov para garantir continuidade da medida probabilidade.

(K5) Continuidade Monotônica. Se para todo $n > 0$, $A_{n+1} \subseteq A_n$ e $\cap_n A_n = \emptyset$, então

$$\lim_{n\to\infty} P(A_n) = 0.^6$$

Um forma equivalente de (K5) é a seguinte, que, conforme visto anteriormente, também é uma propriedade da frequência relativa:

(K5)′ σ-aditividade. Se $\{A_n\}$ é uma coleção enumerável de eventos disjuntos dois a dois, então

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

Teorema 1.8.1: *Se P satisfaz (K1)—(K4), então P satisfaz (K5)′ se, e somente se, satisfaz (K5).*

Prova: Primeiro, será provado que (K1)—(K5) implicam o axioma da σ-aditividade (K5)′. Seja $\{A_i\}$ qualquer sequência enumerável de eventos disjuntos dois a dois, e defina para todo n

$$B_n = \cup_{i>n} A_i,$$

$$\cup_{i=1}^{\infty} A_i = B_n \cup (\cup_{i=1}^{n} A_i).$$

Claramente, para todo $i \leq n$, tem-se que A_i e B_n são disjuntos. Por (K4), tem-se

$$P(\cup_{i=1}^{\infty} A_i) = P(B_n) + \sum_{i=1}^{n} P(A_i).$$

Por definição de série numérica,

$$\lim_n \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{\infty} P(A_i).$$

(K5)′ segue-se se se mostrar que $\lim_n P(B_n) = 0$. Note que $B_{n+1} \subseteq B_n$, e que $\cap_{n=1}^{\infty} B_n = \emptyset$. Então por (K5), o limite acima é zero e K4′ é verdadeiro.

Agora, será provado que (K1)—(K4), (K5)′ implicam o axioma da continuidade monotônica (K5). Seja $\{B_n\}$ qualquer coleção enumerável de eventos satisfazendo as hipóteses do axioma (K5): $B_{n+1} \subseteq B_n$ e $\cap_{n=1}^{\infty} B_n = \emptyset$. Definindo, $A_n = B_n - B_{n+1}$ observa-se que $\{A_n\}$ é uma coleção enumerável de eventos disjuntos dois a dois e que

$$B_n = \cup_{j\geq n} A_j.$$

[6] (K5) (ou equivalentemente (K5)′) é uma idealização que não é aceita por alguns tratamentos subjetivistas de probabilidade, em especial não é aceita por uma escola de estatísticos liderados por deFinetti (1972). Assumir apenas aditividade finita, embora pareça mais plausível, pode levar a complicações inesperadas em teoria estatística. Portanto, neste livro, prossegue-se sob a suposição que o axioma da continuidade (K5) é válido.

Então, por (K5)′,

$$P(B_n) = P(\cup_{j\geq n} A_j) = \sum_{j\geq n} P(A_j).$$

Como por (K5)′,

$$\sum_{j=1}^{\infty} P(A_j) = P(\cup_{j=1}^{\infty} A_j) \leq 1,$$

então

$$\lim_n P(B_n) = \lim_n \sum_{j\geq n} P(A_j) = 0,$$

logo (K5) é verdadeiro. ∎

Definição 1.8.2: Uma função que satisfaz (K1)—(K5) é chamada de uma *medida de probabilidade*.

A terna $(\Omega, \mathcal{A}, P)$ é chamada de **espaço de probabilidade**. Intuitivamente quando se modela uma problema através de probabilidade, basicamente, o que se faz é especificar cada uma das componentes da terna acima.

Eventos são os elementos de $\mathcal{A}$, aos quais se pode atribuir probabilidade. Probabilidade é uma função cujo argumento é um conjunto. Portanto, não somente conjuntos, como também as operações sobre eles, têm uma importância fundamental em teoria da probabilidade. Entretanto, é preciso que a linguagem de conjuntos seja traduzida para a linguagem de probabilidade. A Tabela 1.1 exibe algumas dessas traduções. A idéia subjacente é que um experimento aleatório foi realizado e aconteceu algum evento.

Tabela 1.1: Interpretações interessantes.

Ω	conjunto universo	espaço amostral, evento certo
ω	elemento	evento elementar ou resultado do experimento
A	conjunto A	evento A
$\emptyset$	conjunto vazio	evento impossível
A^c ou $\overline{A}$	complemento de A	não ocorreu o evento A
$A \cap B$	A intersecção B	os eventos A e B ocorreram
$A \cup B$	A união B	os eventos A ou B ocorreram
$\cap_n A_n$	intersecção dos conjuntos A_n	todos os eventos A_n ocorreram
$\cup_n A_n$	união dos conjuntos A_n	ao menos um dos eventos A_n ocorreu

1.8.1 Exemplos de Medidas de Probabilidade

Probabilidade clássica Assumindo que Ω é finito e que os resultados do experimento são equiprováveis, $\forall A \subseteq \Omega$ define-se a probabilidade clássica por

$$P(A) = \frac{||A||}{||\Omega||}, \tag{1.1}$$

onde $||\Omega||$ é o número de resultados *possíveis* (número de elementos do espaço amostral) e $||A||$ é o número de resultados *favoráveis* a A (número de elementos de A) dentre o número de resultados possíveis. definido para qualquer subconjunto A de Ω. O fato que $0 \leq ||A|| \leq ||\Omega||$ e que

$$||A \cup B|| = ||A|| + ||B|| - ||A \cap B||,$$

permitem verificar que P satisfaz os axiomas de Kolmogorov.

A definição pode ser aplicada apenas a uma classe limitada de problemas, isto é, aqueles onde é possível contar os elementos do espaço amostral, Ω, e do evento A. Nessa contagem a técnica usada é Análise Combinatória, que será estudada com mais detalhes no próximo capítulo.

O exemplo a seguir [7] calcula probabilidades usando (1.1). Adicionalmente, a expressão (1.2) mostra que, neste caso, existe uma fórmula fechada,

$$\log_b(1 + 1/k),$$

para o cálculo de

$$\lim_{N\to\infty} N(k)/N.$$

Exemplo 1.8.3: Todo número real x é unicamente representado na expansão b-ádica (Kulisch & Miranker, 1981)

$$x = *d_n d_{n-1} \ldots d_1 d_0 . d_{-1} d_{-2} \ldots = \sum_{i=n}^{-\infty} d_i b^i,$$

onde $* \in \{+, -\}$ é o sinal do número, b é a base da representação, $b \in \mathbb{N}$, $b > 1$, d_i, $i = n, \ldots, -\infty$, são inteiros positivos tais que $0 \leq d_i \leq b - 1$ e $d_i \leq b - 2$ para infinitamente muitos i.

Sejam a, b, k, n, N inteiros positivos tais que $a, b, N \geq 2$, $k = 1, \cdots, b - 1$, $n = 1, \cdots, N$. Seja $N(k)$ o número de vezes que k aparece como o primeiro dígito de $\{a^n\}_{n=1}^{N}$ na base b. Sabe-se que

$$\lim_{N\to\infty} N(k)/N = \log_b(1 + 1/k). \tag{1.2}$$

As Tabelas 1.2 e 1.3 apresentam resultados computacionais para k, $N(k)$ e

$$P(k, N) = N(k)/N,$$

que é a distribuição de probabilidade clássica do primeiro dígito significativo, onde $b = 10$, $N = 10^2, 10^3, 10^4, 10^5$ $a = 2, 3$.

[7] R. A. Mendoza e M. A. Campos, *The Limit Probability Distribution of the First Digit of a^n*, manuscrito não publicado baseado em R. Raimi (1976) *The first digit problem*, *Amer. Math. Monthly* 83, pp 521-538.

Tabela 1.2: k, $N(k)$ e $P(k,N)$ para 2^n, $n = 1, \cdots, N$ e $N = 10^2, 10^3, 10^4, 10^5$.

k	$N(k)$	$P(k,10^2)$	$N(k)$	$P(k,10^3)$	$N(k)$	$P(k,10^4)$	$N(k)$	$P(k,10^5)$
1	30	0.30	301	0.301	3010	0.3010	30103	0.30103
2	17	0.17	176	0.176	1761	0.1761	17611	0.17611
3	13	0.13	125	0.125	1249	0.1249	12492	0.12492
4	10	0.10	97	0.097	970	0.0970	9692	0.09692
5	7	0.07	79	0.079	791	0.0791	7919	0.07919
6	7	0.07	69	0.069	670	0.0670	6695	0.06695
7	6	0.06	56	0.056	579	0.0579	5797	0.05797
8	5	0.05	52	0.052	512	0.0512	5116	0.05116
9	6	0.05	45	0.045	458	0.0458	4576	0.04576

Tabela 1.3: k, $N(k)$ e $P(k,N)$ para 3^n, $n = 1, \cdots, N$ e $N = 10^2, 10^3, 10^4, 10^5$.

k	$N(k)$	$P(k,10^2)$	$N(k)$	$P(k,10^3)$	$N(k)$	$P(k,10^4)$	$N(k)$	$P(k,10^5)$
1	28	0.28	300	0.300	3007	0.3007	30101	0.30101
2	19	0.19	177	0.177	1764	0.1764	17611	0.17611
3	12	0.12	123	0.123	1247	0.1247	12492	0.12492
4	8	0.08	98	0.098	968	0.0968	9693	0.09693
5	9	0.09	79	0.079	792	0.0792	7916	0.07916
6	7	0.07	66	0.066	669	0.0669	6697	0.06697
7	7	0.07	59	0.059	582	0.0582	5798	0.05798
8	5	0.05	52	0.052	513	0.0513	5116	0.05116
9	5	0.05	46	0.046	458	0.0458	4576	0.04576

A Tabela 1.4 exibe valores numéricos aproximados para o resultado teórico

$$\log_b(1 + 1/k),$$

quando $a = 2$ e $N = 10^5$.

Tabela 1.4: Valores para $\log_b(1+1/k)$.

k	$\log_{10}(1+1/k)$
1	0.30103
2	0.17609
3	0.12385
4	0.09691
5	0.07918
6	0.06818
7	0.05690
8	0.05115
9	0.04532

■

Probabilidade frequentista

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n},$$

onde n_A é o número de ocorrências de A em n ensaios independentes do experimento (teoria baseada na observação).

O problema quando da aplicação desta definição para calcular a probabilidade de um evento é: quando é que n é suficientemente grande, isto é, quando é que o experimento aleatório foi realizado um número suficientemente grande de vezes, para garantir que a frequência relativa do evento A é $P(A)$? A resposta formal a esta pergunta será respondida no estudo de teoremas limite.

Exemplo 1.8.4: Simule o lançamento de uma moeda para constatar que quando uma moeda é lançada um número grande de vezes as probabilidades de cara e coroa tornam-se aproximadamente as mesmas. ■

Probabilidade geométrica. Considerando o espaço amostral constituído de objetos geométricos tais como pontos, retas e planos, a obtenção de probabilidades, nesse caso, é referenciada na literatura como problemas de probabilidade geométrica. Portanto, dado um certo evento A, nesse contexto, de modo geral,

$$P(A) = \frac{m(A)}{m(\Omega)},$$

desde que todas as medidas estejam bem definidas.

Por exemplo, suponha que um ponto seja escolhido aleatoriamente no quadrado $0 \leq x \leq 1$, $0 \leq y \leq 1$. Pode-se encontrar a probabilidade de que o ponto pertença à região limitada pelas retas $x \geq 1/2$ e $x + y \geq 1/3$, através da razão entre a área desta região, que é $1/2$, pela área do quadrado $0 \leq x \leq 1$, $0 \leq y \leq 1$, que é 1. Logo a probabilidade é igual a $1/2$.

Espaço amostral enumerável. O número de elementos de Ω é finito, mas os eventos elementares não são necessariamente equiprováveis. Seja $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ um conjunto finito, e seja $P(\{\omega_i\}) = p_i$, onde $p_i \geq 0$, $i = 1, \cdots n$ e $\sum_{i=1}^n p_i = 1$, e $P(A) = \sum_{\omega_i \in A} P(\{\omega_i\})$. Neste caso, também é fácil verificar que P é uma medida de probabilidade, verificando os axiomas de Kolmogorov.

1.8.2 Propriedades de uma Medida Probabilidade

Teorema 1.8.5: *Se P é uma medida de probabilidade, então*

(i) $P(A^c) = 1 - P(A)$.

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

(iii) $P(A) \leq 1$.

(iv) Monotonicidade. *Se $A \subseteq B$, então $P(A) \leq P(B)$.*

(v) $A_1 \subset A_2 \Rightarrow P(A_2 - A_1) = P(A_2) - P(A_1)$.

(vi) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

(vii) *Princípio da Inclusão-Exclusão*

$$\begin{aligned} P(\cup_{i=1}^n A_i) &= \sum_i P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) \\ &= + \ldots + (-1)^{n-1} P(A_1 \cap A_2 \cap \ldots \cap A_n). \end{aligned}$$

(viii) $P(A \cup B) \geq \max\{P(A), P(B)\} \geq \min\{P(A), P(B)\} \geq P(A \cap B)$.

(ix) *Sejam $A_1 \subset A_2 \subset \ldots$, tal que $\cup_{n=1}^\infty A_n = A$, então $\lim_{n\to\infty} P(A_n) = P(A)$. (continuidade da probabilidade)*

(x) *Sejam $A_1 \supset A_2 \supset \ldots$, tal que $\cap_{n=1}^\infty A_n = A$, então $\lim_{n\to\infty} P(A_n) = P(A)$. (continuidade da probabilidade)*

Prova:

(i) Segue-se do fato que $\Omega = A \cup A^c$, (K3), e (K4), pois

$$1 = P(\Omega) = P(A) + P(A^c).$$

(ii) $\Omega^c = \emptyset$, e por (i)

$$P(\emptyset) = 1 - P(\Omega) = 0.$$

(iii) $1 = P(\Omega) = P(A) + P(A^c) \geq P(A)$, visto que $P(A^c) \geq 0$ por (K2).

(iv) $B = A \cup (B - A)$, onde A e $B - A$ são disjuntos. Então (K4) implica que $P(B) = P(A) + P(B - A)$. O resultado segue do fato que $P(B - A) \geq 0$ (K2).

(v)
$$A_1 \subset A_2 \Rightarrow A_2 = A_1 \cup (A_2 \cap A_1^c) \Rightarrow P(A_2) = P(A_1) + P(A_2 \cap A_1^c).$$

Como $A_2 \cap A_1^c = A_2 - A_1$, o resultado segue-se.

(vi) $A \cup B = A \cup (B - A)$, e A e $B - A$ são disjuntos, (K4) implica que $P(A \cup B) = P(A) + P(B - A)$; como $B = (A \cap B) \cup (B - A)$, onde $A \cap B$ e $B - A$ são disjuntos, (K4) implica que $P(B) = P(A \cap B) + P(B - A)$. Logo,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

(vii) Esta prova é por indução. Seja $B = A_1 \cup A_2 \cup \ldots A_{n-1}$. Logo, $\cup_{i=1}^n A_i = B \cup A_n$, e portanto, pela propriedade anterior,

$$P(\cup_{i=1}^n A_i) = P(B) + P(A_n) - P(B \cap A_n).$$

Mas,

$$\begin{aligned} B \cap A_n &= (A_1 \cup A_2 \cup \ldots \cup A_{n-1}) \cap A_n \\ &= (A_1 \cap A_n) \cup (A_2 \cap A_n) \cup \ldots \cup (A_{n-1} \cap A_n). \end{aligned}$$

Logo,

$$\begin{aligned} P(B \cap A_n) &= P(A_1 \cap A_n) + \ldots + P(A_{n-1} \cap A_n) \\ &= -P((A_1 \cap A_n) \cap (A_2 \cap A_n)) - \ldots + (-1)^{n-2} P(A_1 \cap A_2 \ldots A_n). \end{aligned}$$

Lembrando que $P(B) = P(A_1 \cup A_2 \cup \ldots \cup A_{n-1})$, aplicando a hipótese indutiva e realizando as simplificações entre conjuntos, o resultado está provado.

No caso particular de $n = 3$, o princípio de inclusão-exclusão afirma que

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &\quad -P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\ &\quad P(A_1 \cap A_2 \cap A_3). \end{aligned}$$

(viii) Sem perda de generalidade, sejam

$$P(A) = \min\{P(A), P(B)\}$$

e

$$P(B) = \max\{P(A), P(B)\}.$$

Como $B \subseteq A \cup B \Rightarrow P(B) \leq P(A \cup B) \Rightarrow$

$$P(A \cup B) \geq \max\{P(A), P(B)\}.$$

Obviamente,

$$\max\{P(A), P(B)\} \geq \min\{P(A), P(B)\}.$$

De $A \cap B \subseteq A$, tem-se que $P(A \cap B) \leq P(A)$. Logo,

$$\min\{P(A), P(B)\} \geq P(A \cap B).$$

(ix) Construindo uma sequência, $\{B_n\}$, de elementos excludentes:

$$B_1 = A_1$$

$$B_2 = A_2 \cap A_1^c$$

$$\cdots$$

$$B_n = A_n \cap A_{n-1}^c$$

$$\cdots$$

Tem-se que:

$$\cup_{n=1}^{\infty} A_n = A = \cup_{n=1}^{\infty} B_n$$

e

$$A_n = \cup_{k=1}^{n} B_k.$$

Logo,

$$\begin{aligned} \lim_{n\to\infty} P(A_n) &= \lim_{n\to\infty} P(\cup_{k=1}^{n} B_k) \\ &= P(\cup_{n=1}^{\infty} B_n) \\ &= P(\cup_{n=1}^{\infty} A_n) \\ &= P(A). \end{aligned}$$

(x) Como $A_n \supset A_{n+1}$, $\forall n \geq 1$, então $A_n^c \subset A_{n+1}^c$. Do item anterior tem-se que

$$\lim_{n\to\infty} P(A_n^c) = P(\cup^{\infty} A_n^c) = P(A^c).$$

Logo,

$$\begin{aligned}\lim_{n\to\infty} P(A_n) &= \lim_{n\to\infty}(1-P(A_n^c))\\ &= 1-\lim_{n\to\infty} P(A_n^c)\\ &= 1-P(A^c)\\ &= P(A).\end{aligned}$$

▮

A notação usada neste capítulo é a comumente encontrada nos livros de probabilidade. Entretanto, fora do contexto de probabilidade, é possível, aliás quase certo, encontrar notação distinta. Por exemplo, em Russel & Norvig (1995) tem-se $P(A \vee B)$, $P(A \wedge B)$ e $P(\neg A)$ para $P(A \cup B)$, $P(A \cap B)$, $P(A^c)$.

Teorema 1.8.6: Probabilidade de Partições. *Se $\{A_i\}$ é uma partição enumerável (ou finita) de Ω composta de conjuntos em $\mathcal{A}$, então para todo $B \in \mathcal{A}$*

$$P(B) = \sum_i P(B \cap A_i).$$

Prova: Como $\{A_i\}$ é uma partição, segue-se que

$$B = B \cap \Omega = B \cap (\cup_i A_i) = \cup_i (B \cap A_i).$$

O resultado segue vem por (K5)$'$. ▮

Teorema 1.8.7: Desigualdade de Boole. *Para n eventos arbitrários $\{A_1, \ldots, A_n\}$, a desigualdade de Boole é*

$$P(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i).$$

Prova: Seja $n = 2$. Logo, $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2)$ porque $P(A_1 \cap A_2) \geq 0$. Usar indução para provar para n. ▮

Corolário 1.8.8: *Para n eventos arbitrários $\{A_1, \ldots, A_n\}$,*

$$P(\cap A_i) \geq \sum_{i=1}^n P(A_i) - (n-1).$$

Prova: Utilizando a Lei de De Morgan e a desigualdade de Boole para os eventos $\{A_1^c, \ldots, A_n^c\}$,

$$P(\cup_{i=1}^n A_i^c) = 1 - P(\cap_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i^c) = \sum_{i=1}^n (1 - P(A_i)).$$

Logo,

$$P(\cap_{i=1} n A_i) \geq \sum_{i=1}^n P(A_i) - (n-1).$$

▮

Exemplo 1.8.9: Em um grupo de r pessoas qual a probabilidade de haver pelo menos duas pessoas que completem aniversário no mesmo dia, assumindo que a distribuição de aniversários é uniforme ao longo do ano e desprezando a existência de anos bissextos?
Solução: Para determinar esta probabilidade a probabilidade usada é a clássica. O número de resultados possíveis para os aniversários de r pessoas é 365^r. O número de casos possíveis onde todas as pessoas fazem aniversário em dias diferentes é dado por $365 \times 364 \times \cdots \times (365-(r-1))$. Portanto, o número de casos possíveis onde pelo menos duas pessoas fazem aniversário no mesmo dia é a diferença entre o número total de aniversários possíveis e o nùmero de casos onde as pessoas tém aniversários em datas diferentes, ou seja, é igual a

$$365^r - 365 \times 364 \times \cdots \times (365-(r-1)).$$

Logo, a probabilidade deste evento é:

$$1 - \frac{365 \times 364 \times \cdots \times (365-(r-1))}{365^r}.$$

Para $r = 23$, essa probabilidade é aproximadamente igual a 0.51. E para $r = 50$, 0.97. ■

Exemplo 1.8.10: Em uma loteria de N números há um só prêmio. Salvador compra n $(1 < n < N)$ bilhetes para uma só extração e Sílvio compra n bilhetes, um para cada uma de n extrações. Qual dos dois jogadores têm mais chance de ganhar algum prêmio?
Solução: A probabilidade de Salvador ganhar algum prêmio é $\frac{n}{N}$. O número total de n extrações possíveis é N^n. O número de casos onde Sílvio não ganha qualquer prêmio é $(N-1)^n$, logo, o número de casos onde Sílvio ganha algum prêmio é igual a $N^n - (N-1)^n$. Portanto, a probabilidade de Sílvio ganhar algum prêmio é $1 - \frac{(N-1)^n}{N^n}$.

Por indução prova-se que Salvador tem mais chance de ganhar, ou seja, $\frac{n}{N} > 1 - \frac{(N-1)^n}{N^n}$, que equivale a

$$\frac{(N-1)^n}{N^n} > 1 - \frac{n}{N}.$$

Para $n = 2$:

$$\frac{(N-1)^2}{N^2} = 1 - \frac{2}{N} + \frac{1}{N^2} > 1 - \frac{2}{N}.$$

Suponha que para $n = k$,

$$\frac{(N-1)^k}{N^k} > 1 - \frac{k}{N}.$$

Multiplicando esta expressão por $\frac{N-1}{N}$,

$$\frac{(N-1)^{k+1}}{N^{k+1}} > (\frac{N-1}{N})(1 - \frac{k}{N}) = 1 - \frac{1}{N} - \frac{k}{N} + \frac{k}{N^2} > 1 - \frac{k+1}{N}.$$

■

Exemplo 1.8.11: Doze pessoas são divididas em três grupos de 4. Qual é a probabilidade de duas determinadas dessas pessoas ficarem no mesmo grupo?

Solução: O número total de divisões de doze pessoas em 3 grupos de 4 é $\binom{12}{4}\binom{8}{4}\binom{4}{4}$. Para contar o número de casos favoráveis ao evento, sabe-se que existem 3 opções de escolha sobre em qual grupo as duas pessoas determinadas podem ficar. Das 10 pessoas restantes, tem de se escolher mais duas para estarem neste grupo, o que pode ser resolvido de $\binom{10}{2}$ maneiras diferentes. E $\binom{8}{4}\binom{4}{4}$ são maneiras diferentes de dividir as outras 8 pessoas nos dois grupos restantes. Portanto, a probabilidade de duas determinadas pessoas ficarem no mesmo grupo é:

$$\frac{3\binom{10}{2}\binom{8}{4}\binom{4}{4}}{\binom{12}{4}\binom{8}{4}\binom{4}{4}} = \frac{3}{11}.$$

■

Exemplo 1.8.12: Suponha que numa sala estão n mães cada uma com um filho. Suponha que duplas sejam formadas aleatoriamente, onde cada dupla contém uma mãe e um filho. Qual éa probabilidade de que pelo menos uma mãe forme uma dupla com seu próprio filho?
Solução: Seja A_i o evento que a i-ésima mãe forma dupla com seu filho. O objetivo é determinar

$$P(\cup_{i=1}^{n} A_i).$$

Calculando esta probabilidade utilizando a fórmula da inclusão-exclusão. Note que:

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n} \text{ para todo } i \in \{1, 2, \ldots, n\}$$
$$P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} \text{ para } i \neq j$$

e em geral, para um grupo $I \in \{1, 2, \ldots, n\}$ de mães,

$$P(\cap_{i\in I} A_i) = \frac{(n - ||I||)!}{n!}.$$

Como existem $\binom{n}{||I||}$ grupos de mães com cardinalidade $||I||$,

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{i} \frac{(n-i)!}{n!}$$
$$= \sum_{i=1}^{n} (-1)^{i+1} \frac{1}{i!}$$

Note que quando $n \to \infty$, esta probabilidade tende a $1 - \frac{1}{e}$. ■

Exemplo 1.8.13: Demonstre que se $P(A_i) = 1$ para $i = 1, 2, \ldots$, então $P(\cap_{i=1}^{\infty} A_i) = 1$.
Solução: Como $P(A_i) = 1$, tem-se que $P(A_i^c) = 1 - P(A_i) = 0$. Logo, pela desigualdade de Boole, $P(\cup_{i=1}^{\infty} A_i^c) \leq \sum_{i=1}^{\infty} P(A_i^c) = 0$. Portanto, $P(\cup_{i=1}^{\infty} A_i^c) = 0$ e pela Lei de De'Morgan, $\cap_{i=1}^{\infty} A_i = (\cup_{i=1}^{\infty} A_i^c)^c$, tem-se que $P(\cap_{i=1}^{\infty} A_i) = 1 - P(\cup_{i=1}^{\infty} A_i^c) = 1$. ■

Exemplo 1.8.14: Demonstre: se $A_1, A_2, \ldots$ e $B_1, B_2, \ldots$ são eventos do mesmo espaço de probabilidade tais que $P(A_n) \to 1$ e $P(B_n) \to p$, então $P(A_n \cap B_n) \to p$.
Solução: Note que

$$\begin{aligned} & P(A_n \cap B_n) = 1 - P((A_n \cap B_n)^c) = 1 - P(A_n^c \cup B_n^c) \\ & \geq 1 - P(A_n^c) - P(B_n^c) = P(A_n) + P(B_n) - 1. \end{aligned} \tag{1.1}$$

Como $P(A_n) + P(B_n) - 1 \to p$, tem-se que $\liminf P(A_n \cap B_n) \geq p$. Por outro lado, como $P(A_n \cap B_n) \leq P(B_n)$ e $P(B_n) \to p$, tem-se que $\limsup P(A_n \cap B_n) \leq p$. Portanto, $\lim P(A_n \cap B_n) = p$. ■

1.9 Aprendendo um pouco mais

Teorema 1.9.1: *Se (A_n) é uma sequência de suconjuntos de um conjunto Ω tal que $A_1 \subseteq A_2 \subseteq A_3 \ldots$, então $\lim A_n = \cup_{n=1}^{\infty} A_n$.*

Teorema 1.9.2: *Se (A_n) é uma sequência de suconjuntos de um conjunto Ω tal que $A_1 \supseteq A_2 \supseteq A_3 \ldots$, então $\lim A_n = \cap_{n=1}^{\infty} A_n$.*

Dado estes resultados, os itens (ix) e (x) do Teorema 1.8.5 nos permitem afirmar que para sequências monotônicas de eventos o limite da sequência comuta com a probabilidade, isto é,

$$P(\lim A_n) = \lim P(A_n).$$

Teorema 1.9.3: *O conjunto dos números reais é não-enumerável.*

Se o espaço amostral for finito, toda álgebra é uma σ-álgebra, pois só existe um número finito de eventos distintos. Se o espaço amostral for infinito, existem álgebras que não são σ-álgebras, como mostra o exemplo seguinte.

Exemplo 1.9.4: Um conjunto é co-finito se seu complementar for finito. A coleção de conjuntos de números reais finitos e co-finitos é uma álgebra que não é uma σ-álgebra.

Lema 1.9.5: *Se $\mathcal{A}$ é uma σ-álgebra, então $\Omega \in \mathcal{A}$*

Prova: Como $\mathcal{A}$ é não vazio, seja A um seu elemento qualquer. Pela segunda propriedade de álgebras, tem-se que $A^c \in \mathcal{A}$, e pela terceira, $\Omega = A \cup A^c \in \mathcal{A}$. ▮

Teorema 1.9.6: *Sejam $\mathcal{A}_1$ e $\mathcal{A}_2$ álgebras (σ-álgebras) de subconjuntos de Ω e seja $\mathcal{A} = \mathcal{A}_1 \cap \mathcal{A}_2$ a coleção de subconjuntos comuns às duas álgebras. Então, $\mathcal{A}$ úma álgebra (σ-álgebra).*

Prova: Como $\mathcal{A}_1$ e $\mathcal{A}_2$ são álgebras, ambos contém Ω. Então, $\Omega \in \mathcal{A}$. Se $A \in \mathcal{A}$, então A está em ambos $\mathcal{A}_1$ e $\mathcal{A}_2$. Logo, A^c está em ambos $\mathcal{A}_1$ e $\mathcal{A}_2$, e portanto na sua intersecção $\mathcal{A}$. Se $A, B \in \mathcal{A}$, então eles estão em ambos $\mathcal{A}_1$ e $\mathcal{A}_2$. Consequentemente, $A \cup B$ está em ambos $\mathcal{A}_1$ e $\mathcal{A}_2$ e, portanto, em $\mathcal{A}$. Como $\mathcal{A}$ satisfaz as três condições da definição de álgebra de eventos, $\mathcal{A}$ é uma álgebra de eventos. A prova no caso de σ-álgebras é análoga. ▮

Corolário 1.9.7: *Existe uma menor (no sentido de inclusão) álgebra (σ-álgebra) contendo qualquer família dada de subconjuntos de Ω.*

Prova: Seja $\mathcal{C}$ uma coleção qualquer de subconjuntos de Ω. Defina $\mathcal{A}(\mathcal{C})$ como sendo o conjunto que é igual a intersecção de todas as álgebras de eventos que contém $\mathcal{C}$, isto é:

$$\mathcal{A}(\mathcal{C}) = \bigcap_{\mathcal{A}\supseteq\mathcal{C}:\mathcal{A} \text{ é uma álgebra de eventos}} \mathcal{A}.$$

Pelo Teorema 1.9.6, $\mathcal{A}(\mathcal{C})$ é uma álgebra de eventos, e consequentemente é a menor álgebra de eventos contendo $\mathcal{C}$. A prova no caso de σ-álgebras é análoga. ∎

Seja Ω é um conjunto não-vazio. Se Ω for um espaço topológico, isto é, Ω contiver conjuntos abertos, a nova σ-álgebra sobre a coleção dos abertos de Ω é denominada σ-álgebra de Borel. Se $\Omega = \mathbb{R}$, esta σ-álgebra é denotada por $\mathcal{B}$. Os elementos de $\mathcal{B}$ são chamados conjuntos de Borel, ou borelianos. Os conjuntos $[x, y]$, $(x, y]$ e $[x, y)$ são conjuntos de Borel, pois:

$$[x, y] = \cap_{n=1}^{\infty}(x - \frac{1}{n}, y + \frac{1}{n}),$$

$$[x, y) = \cap_{n=1}^{\infty}(x - \frac{1}{n}, y),$$

$$(x, y] = \cap_{n=1}^{\infty}(x, y + \frac{1}{n}).$$

A σ-álgebra de Borel $\mathcal{B}$ de subconjuntos de reais é, por definição, a menor σ-álgebra contendo todos os intervalos e é a σ-álgebra usual quando se lida com quantidades reais ou vetoriais. Em particular, tem-se que uniões enumeráveis de intervalos (por exemplo, o conjunto dos números racionais), seus complementos (por exemplo, o conjunto dos números irracionais), e muito mais estão em $\mathcal{B}$. Para todos os fins práticos, pode-se considerar que $\mathcal{B}$ contém todos os subconjuntos de reais que consegue-se descrever.

1.10 Exercícios

1.1 **(a)** Uma caixa com 6 chips contém 2 defeituosos. Descreva um espaço amostral para cada uma das situações abaixo:

(a1) Os chips são examinados um a um sem reposição até que um defeituoso seja encontrado.

(a2) Os chips são examinados um a um sem reposição até que todos os defeituosos sejam encontrados.

(b) Generalize o problema. Responda às mesmas questões anteriores, supondo que se tem N chips na caixa, dos quais $n < N$ são defeituosos.

1.2 Coloque V ou F nas sentenças abaixo:

(a) $A = \emptyset \Rightarrow P(A) = 0$. ( )

(b) $P(A) = 0 \Rightarrow A = \emptyset$. ( )

(c) $A = \emptyset \Leftrightarrow P(A) = 0$. ( )

(d) $A \supseteq B \Rightarrow P(A) \leq P(B)$. ( )

(e) $A \subseteq B \Rightarrow P(A) \geq P(B)$. ( )

(f) $A \subseteq B \Rightarrow P(A) \leq P(B)$. ( )

(g) A e B excludentes $\Rightarrow P(A \cup B) = P(A) + P(B)$. ( )

(h) A e B excludentes $\Rightarrow P(A \cap B) = P(A)P(B)$. ( )

1.3 Professor Leônidas está tentando calcular a probabilidade $p = P(A)$ do evento A, e determinou que ela é uma raiz do seguinte polinômio de grau cinco:

$$(p-3)(p-3\sqrt{-1})(p+3\sqrt{-1})(p+0.3)(p-0.3) = 0.$$

Baseado nesta fato, qual é o valor de p?

1.4 Se $\Omega = \{a, b, c\}$, a álgebra $\mathcal{A}$ é o conjunto das partes de Ω e a medida de probabilidade P é parcialmente definida por

$$P(\{a,b\}) = 0.5,\ P(\{b,c\}) = 0.8,\ P(\{a,c\}) = 0.7,$$

então complete a especificação de P para todos os eventos em $\mathcal{A}$.

1.5 Se $\{A_i\}$ for uma partição enumerável de Ω e $P(A_i) = ab^i$, $i \geq 1$, quais as condições que a e b devem satisfazer para que P seja uma medida de probabilidade?

1.6 As seguintes questões não estão relacionadas umas com as outras.

(a) Se $I_A I_B$ for identicamente igual a zero, o que dizer a respeito da relação entre A e B?

(b) Se $A \cap B^c = B \cap A^c$, o que dizer a respeito da relação entre A e B?

(c) Se $I_A^2 + I_B^2$ for identicamente igual a 1, o que concluir sobre A e B?

1.7 Determine se cada uma das afirmações a seguir são verdadeiras ou falsas. Se a relação for falsa, apresente um contra-exemplo. Se for verdadeira, prove-a.

(a) Se $x \in A$ e $A \subset B$, então $x \in B$.

(b) Se $A \subseteq B$ e $B \subseteq C$, então $A \subseteq C$.

(c) Se $A \not\subseteq B$ e $B \subseteq C$, então $A \not\subseteq C$.

(d) Se $A \not\subseteq B$ e $B \not\subseteq C$, então $A \not\subseteq C$.

(e) Se $x \in A$ e $A \not\subseteq B$, então $x \notin B$.

(f) Se $A \subseteq B$ e $x \notin B$, então $x \notin A$.

1.8 Descreva um espaço amostral para cada um dos experimentos abaixo.

(a) *Strings* de dígitos binários são geradas até que pela primeira vez o mesmo resultado apareça duas vezes em sucessão.

(b) *Strings* de dígitos binários são geradas até que o dígito 1 apareça pela primeira vez.

(c) *Strings* de 3 dígitos binários são geradas. Observe as sequências de zeros e uns.

(d) Conte o número de zeros em uma *string* de dígitos binários com n dígitos.

1.9 Mostre que $P(E \cap F) \leq P(E) \leq P(E \cup F) \leq P(E) + P(F)$.

1.10 Um ponto é escolhido ao acaso sobre um quadrado unitário. Determine a probabilidade de que o ponto esteja no triângulo limitado por $x = 0$, $y = 0$ e $x + y = 1$.

1.11 Um ponto é escolhido ao acaso sobre um disco unitário. Determine a probabilidade de que o ponto esteja no setor angular de 0 a $\pi/4$.

1.12 Suponha que A, B e C sejam eventos tais que $P(A) = P(B) = P(C) = 1/4$, $P(A \cap B) = P(B \cap C) = 0$ e $P(A \cap C) = 1/8$. Calcule a probabilidade de que ao menos um dos eventos A, B ou C ocorra.

1.13 Suponha a declaração `if` B `then` s_1 `else` s_2 e que o experimento aleatório consista em observar duas execuções sucessivas desta declaração. Sejam os eventos

$$E_1 = \{\text{pelo menos uma execução de } s_1\}$$

e

$$E_2 = \{\text{a declaração } s_2 \text{ é executada pela primeira vez}\}.$$

(a) Exiba um espaço amostral para o experimento.

(b) Calcule $P(E_1)$ e $P(E_2)$ supondo equiprobabilidade onde necessário.

1.14 *Distribuição de Números Primos*

(a) Considere os intervalos $A_k = [10k, 10(k+1))$, $k = 0 \cdots, 9$. Sejam, n o total dos primos em [0,100) e n_k a frequência deles em cada A_k. Seja $p_k = \frac{n_k}{n}$. Calcule p_k e faça um gráfico com os pontos (k, p_k), para $k = 0 \cdots, 9$.

(b) Repita todo o problema anterior com $A_k = [100k, 100(k+1))$, $k = 0 \cdots, 9$, e n o total dos primos em [0,1000).

(c) Agora com $A_k = [1000k, 1000(k+1))$, $k = 0 \cdots, 9$, sendo n o total dos primos em [0,10000).

(d) Os resultados que você obteve, empiricamente, aceitam ou refutam a seguinte afirmação: *números primos ocorrem menos frequentemente entre inteiros maiores que entre inteiros menores*.

(e) Seja $\pi(x)$ o número de primos menores que $x \in \mathbb{R}$, $x > 0$. De acordo com seus cálculos, qual afirmação abaixo você aceita como sendo verdadeira?

$$\pi(x) \geq \lfloor \log_2(\log_2 x) \rfloor + 1,$$

$$\pi(x) \leq \lfloor \log_2(\log_2(x) \rfloor + 1,$$

onde em (a) $x = 100$, em (b) $x = 1000$ e em (c) $x = 10000$.

1.15 Sejam $A_1, A_2, \cdots$, $B_1, B_2, \cdots$ eventos aleatórios definidos no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$. Mostre que:

(a) $P(\cap_{k=1}^n A_k) \geq 1 - \sum_{k=1}^n P(A_k^c)$.

(b) Se $P(A_k) \geq 1 - \varepsilon$ para $k = 1, \cdots, n$, então $P(\cap_{k=1}^n A_k) \geq 1 - n\varepsilon$.

(c) Se $P(A_n) = 0$ para $n = 1, 2, \cdots$ então $P(\cup_{n=1}^\infty A_n) = 0$.

1.16 Para todo conjunto unidimensional A para o qual a integral existe seja $P(A) = \int_A f(x)dx$, onde $f(x) = 6x(1-x)$, $0 < x < 1$ e zero para $x \notin (0,1)$. Se $A_1 = \{x \mid \frac{1}{4} < x < \frac{3}{4}\}$ e $A_2 = \{x \mid x = \frac{1}{2}\}$, calcule $P(A_1)$, $P(A_2)$, $P(A_1 \cap A_2)$, $P(A_1 \cup A_2)$.

1.17 Seja a probabilidade do evento A,

$$P(A) = \int_A e^{-x} dx,\ 0 < x < \infty,$$

e seja $A_k = \{x \mid 2 - 1/k < x \leq 3\}$, $k = 1, 2, \cdots$. Mostre que $\lim_{k\to\infty} P(A_k) = P(\lim_{k\to\infty} A_k)$. Seja agora $A_k = \{x \mid 1/k - 2 < x \leq 3\}$, $k = 1, 2, \cdots$. Mostre que $\lim_{k\to\infty} P(A_k) = P(\lim_{k\to\infty} A_k)$.

1.18 Um poliedro com k faces, $k > 3$, rotuladas $f_1, f_2, \cdots, f_k$ é atirado aleatoriamente em um plano, sendo observada a face tangente ao mesmo.

(a) Descreva o espaço amostral.

(b) Seja o evento A, a face voltada para baixo não excede o número $k/2$. Descreva A.

(c) Calcule $P(A)$ para um icosaedro, dodecaedro e octaedro.

1.19 Uma coleção de 100 programas foi checada com respeito a erros de sintaxe, S, erros de entrada e saída, I, e outros tipos de erros, E. A quantidade de vezes que os erros ocorreram foi: S, 20 vezes; I, 10 vezes; E, 5 vezes; $S \wedge I$, 6 vezes; $S \wedge E$, 3 vezes; $I \wedge E$, 2 vezes; $S \wedge I \wedge E$, 1 vez. Um programa é selecionado aleatoriamente. Calcule a probabilidade de que este apresente

(a) S ou I;

(b) ao menos um tipo de erro.

1.20 Dois dados são lançados. Considere os eventos

$$A = \{\text{a soma dos pontos sobre as duas faces é um número par}\},$$

$$B = \{\text{1 aparece em pelo menos um dos dados}\}.$$

Descreva os eventos: (a) $A \cap B$; (b) $A \cup B$; (c) $A \cap \overline{B}$.

1.21 Um alvo consiste de dez círculos concêntricos com raios $r_k, k = 1, 2, \ldots 10$, onde $r_1 < r_2 < \ldots r_{10}$. O evento A_k indica um acerto no círculo de raio k. Descreva em palavras os eventos $B = \cup_{k=1}^{6} A_k$ e $C = \cap_{k=5}^{10} A_k$.

1.22 Um experimento consiste em se retirar sem reposição 3 impressoras de um lote e testá-las de acordo com alguma característica de interesse. Assinale D, para impressora defeituosa e B, para perfeita. Sejam os eventos:

$$A_1 = \{\text{a 1a. impressora foi defeituosa}\},$$
$$A_2 = \{\text{a 2a. impressora foi defeituosa}\},$$
$$A_3 = \{\text{a 3a. impressora foi defeituosa}\}.$$

(a) Descreva o espaço amostral.

(b) Liste todos os elementos de cada um dos seguintes eventos: A_1, A_2, $A_1 \cup A_2$, $A_2 \cap A_3$, $A_1 \cup A_2 \cup A_3$, $A_1 \cap A_2 \cap A_3$.

(c) Explique, em palavras, o significado dos eventos acima.

1.23 Seja A o evento "pelo menos um entre três itens checados é defeituoso", e B o evento "todos os três itens são bons". Descreva os eventos: (a) $A \cup B$; (b) $A \cap B$; (c) $\overline{A}$; (d) $\overline{B}$.

1.24 Há três edições diferentes cada uma contendo pelo menos três volumes. Os eventos A, B e C, respectivamente indicam que pelo menos um livro é escolhido da primeira, da segunda e da terceira edição. Sejam

$$A_s = \{s \text{ volumes são escolhidos da primeira edição}\},$$

$$B_k = \{k \text{ volumes são escolhidos da segunda edição}\}.$$

Qual é o significado dos eventos: (a) $A \cup B \cup C$; (b) $A \cap B \cap C$; (c) $A \cup B_3$; (d) $A_2 \cup B_2$; (e) $(A_1 \cap B_3) \cup (A_3 \cap B_1)$?

1.25 Um número é escolhido do conjunto dos números naturais. Sejam

$$A = \{\text{o número escolhido é divisível por 5}\}$$

e

$$B = \{\text{o número escolhido termina por 0}\}.$$

Qual é o significado dos eventos $A - B$ e $A \cap \overline{B}$?

1.26 Sejam A, B e C eventos e $A \subset B$. Determine: (a) $A \cap B$; (b) $A \cup B$; (c) $A \cap B \cap C$; (d) $A \cup B \cup C$.

1.27 Mostre que os seguintes eventos formam uma partição do espaço amostral Ω: A, $\overline{A} \cap B$ e $\overline{A \cup B}$.

1.28 Encontre uma condição sob a qual os eventos $A \cup B$, $\overline{A} \cup B$ e $A \cup \overline{B}$ sejam mutuamente exclusivos.

1.29 Suponha que uma instrução leva pelo menos 9 segundos para ser transmitida, processada e a resposta exibida no terminal. O experimento aleatório consiste em mensurar o tempo decorrido da operação completa. Descreva o espaço amostral.

1.30 Uma moeda honesta é lançada até que apareça o mesmo resultados duas vezes seguidas.

(a) Descreva o espaço amostral.

(b) Encontre a probabilidade de que o experimento termine antes de 6 lançamentos.

(c) Encontre a probabilidade de que seja necessário um número par de lançamentos para que o experimento termine.

1.31 Um dado é lançado duas vezes. Seja a o número de pontos obtidos no primeiro lançamento e b o número de pontos obtidos no segundo lançamento. Determine a probabilidade da equação $ax - b = 0$ ter raízes inteiras.

Capítulo 2

Métodos de Contagem

2.1 Introdução

No capítulo anterior foi visto que se $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ é um conjunto finito, então para determinar a probabilidade de qualquer evento A é suficiente especificar a probabilidade de cada evento *simples* ou *elementar* $\{\omega_i\}$, ou seja $P(\{\omega_i\}) = p_i$. É fácil ver que os axiomas de Kolmogorov implicam que $p_i \geq 0, i \geq 1$ e $\sum_{i=1}^{n} p_i = 1$, e $P(A) = \sum_{\omega_i \in A} P(\{\omega_i\})$.

Para se determinar as probabilidades dos eventos simples hipóteses adicionais são necessárias. Por exemplo, se em $\Omega = \{w_1, w_2, w_3\}$, $\{w_1\}$ for 3 vezes mais provável que $\{w_2, w_3\}$, e $\{w_2\}$ for igualmente provável a $\{w_3\}$, tem-se que $p_1 = 3(p_2 + p_3)$, $p_2 = p_3$. Logo, como $p_1 + p_2 + p_3 = 1$ então $p_3 = p_2 = \frac{1}{8}$, e $p_1 = \frac{3}{4}$.

De acordo com a definição clássica de probabilidade onde o espaço amostral Ω é finito e os possíveis resultados do experimento são equiprováveis, a probabilidade de qualquer evento $A \in \mathcal{A}$ é proporcional a sua cardinalidade, isto é,

$$P(A) = \frac{||A||}{||\Omega||}.$$

Portanto, é fundamental contar a quantidade de elementos do evento de interesse quanto do espaço amostral.

Neste capítulo serão estudados métodos de contagem, também conhecidos como métodos de análise combinatória. Embora conjuntos com poucos elementos possam ser contados exaustivamente (força-bruta), conjuntos com cardinalidade moderada podem ser difíceis de contar sem a utilização dessas técnicas matemáticas.

2.2 Regra da Adição

Considere duas atividades 1 e 2 que podem ser realizadas de n_1 e n_2 maneiras distintas, respectivamente. Além disso, suponha que as atividades 1 e 2 sejam excludentes. Então, o número de maneiras pelas quais pode-se executar pelo menos uma das atividades é $n_1 + n_2$.

Generalizando, se existirem k atividades e a i-ésima atividade puder ser realizada de n_i maneiras, $i = 1, 2, \dots, k$, então, o número de maneiras pelas quais pode-se realizar pelo menos uma das atividades é dado por $n_1 + n_2 + \ldots + n_k$, supondo que as atividades sejam excludentes.

Exemplo 2.2.1: Seja o problema de escolher um caminho entre duas cidades A e B dentre três percurssos pelo interior e dois pelo litoral. Portanto existem $3 + 2 = 5$ caminhos disponíveis para a viagem. ■

2.3 Regra da Multiplicação

Considere novamente as atividades 1 e 2 que podem ser realizadas de n_1 e n_2 maneiras, respectivamente. Suponha também que cada maneira de executar 1 possa ser seguida por qualquer maneira para executar 2. Então, a atividade composta formada por 1 seguida de 2 poderá ser executada de $n_1 \times n_2$ maneiras.

Generalizando, se existirem k atividades e a i-ésima atividade puder ser executada de n_i maneiras, $i = 1, 2, \ldots, k$, então a atividade composta formada por 1, seguida por 2,..., seguida pela atividade k, poderá ser executada de $n_1 \times n_2 \times \cdots \times n_k$ maneiras.

Exemplo 2.3.1: Quantos divisores inteiros e positivos possui o número 360? Quantos desses divisores são pares? Quantos são ímpares? Quantos são quadrados perfeitos?
Solução: $360 = 2^3 \times 3^2 \times 5$. Os divisores inteiros e positivos de 360 são os números da forma $2^a \times 3^b \times 5^c$, onde $a \in \{0, 1, 2, 3\}$, $b \in \{0, 1, 2\}$, e $c \in \{0, 1\}$. Portanto, existem $4 \times 3 \times 2 = 24$ maneiras de escolher os expoentes a, b, c. Logo há 24 divisores.

Para o divisor ser par, a não pode ser zero. Então, existem $3 \times 3 \times 2 = 18$ divisores pares. Por outro lado, para o divisor ser ímpar, a tem que ser zero. Logo, existem $1 \times 3 \times 2 = 6$ divisores ímpares. Por fim para o divisor ser quadrado perfeito os expoentes têm que ser pares. Logo, existem $2 \times 2 \times 1 = 4$ divisores quadrados perfeitos. ■

Exemplo 2.3.2: De quantos modos o número 720 pode ser decomposto em um produto de dois inteiros positivos? E o número 144?
Solução: $720 = 2^4 \times 3^2 \times 5$. Os divisores inteiros e positivos de 720 são os números da forma: $2^a \times 3^b \times 5^c$, onde $a \in \{0, 1, 2, 3, 4\}$, $b \in \{0, 1, 2\}$, e $c \in \{0, 1\}$. Portanto, existem $5 \times 3 \times 2 = 30$ maneiras de escolher os expoentes a, b, c. Logo há 30 divisores. Observe que como 720 não é um quadrado perfeito, para cada divisor x de 720 existe um outro divisor $y \neq x$ de 720 tal que $x \times y = 720$. Portanto, cada produto contém dois divisores diferentes de 720. Como existem 30 divisores, existem 15 produtos diferentes.

$144 = 2^4 \times 3^2$. Seguindo o raciocínio anterior, tem-se $5 \times 3 = 15$ divisores de 144. Note que $144 = 12^2$ e este constitui um produto de inteiros positivos que é igual a 144. Os demais produtos contém dois inteiros positivos diferentes que são divisores de 144. Como existem 14 divisores de 144 diferentes de 12, então existem 7 produtos envolvendo estes divisores. Logo, tem-se um total de 8 produtos diferentes. ■

Exemplo 2.3.3: O conjunto A possui 4 elementos e, o conjunto B, 7. Quantas funções $f : A \to B$ existem? Quantas delas são injetoras?
Solução: Para cada elemento de A tem-se 7 possíveis valores diferentes. Como A contém 4 elementos, existem $7 \times 7 \times 7 \times 7 = 7^4$ funções diferentes. Recorde que uma função é injetora

se $f(a) \neq f(b)$ sempre que $a \neq b$. Portanto, o mesmo elemento de B não pode ser imagem de dois elementos de A, logo existem $7 \times 6 \times 5 \times 4 = 840$ funções injetoras. ■

Exemplo 2.3.4: Em uma banca há 5 exemplares iguais da *Veja*, 6 exemplares iguais da *Época* e 4 exemplares iguais da *Isto é*. Quantas coleções não-vazias de revistas dessa banca podem ser formadas?
Solução: Note que cada coleção de revistas vai ser composta por a revistas *Veja*, b revistas Época e c revistas *Isto é*, onde $0 \leq a \leq 5$, $0 \leq b \leq 6$, $0 \leq c \leq 4$, e pelo menos 1 de a, b, ou c é diferente de zero. Então, tem-se $6 \times 7 \times 5 - 1 = 210 - 1 = 209$ diferentes coleções não-vazias dessas revistas. ■

2.4 Amostragem ou escolhas com ou sem reposição

Dado um conjunto com n elementos distintos, o número, $\mu_{n,r}$, de maneiras de selecionar uma sequência distinta de comprimento r escolhida desse conjunto com repetidas seleções do mesmo elemento sendo permitidas, **amostragem com reposição**, é dado por n^r, uma vez que o mesmo procedimento é repetido r vezes e cada procedimento tem n maneiras de ser executado.

Exemplo 2.4.1: Número de Sequências Binárias ou Subconjuntos. O número de sequências binárias de comprimento r é igual a 2^r pois neste caso tem-se que para cada posição i da sequência, $n_i = 2$. O número de subconjuntos de um dado conjunto A, $||A|| = r$, pode ser determinado enumerando $A = \{a_1, a_2, a_3, \ldots, a_r\}$ e descrevendo cada subconjunto B de A por uma sequência binária

$$(b_1, b_2, \ldots, b_r),$$

onde $b_i = 1$ se $a_i \in B$ e $b_i = 0$, caso contrário. Como existem 2^r destas sequências, então existem 2^r subconjuntos de um conjunto de r elementos. Portanto, se $||A|| = r$, o conjunto das partes de A, possui 2^r elementos, o que explica a notação exponencial do conjunto das partes. ■

2.4.1 Arranjos

Dado um conjunto com n elementos distintos, o número $(n)_r$ de maneiras de selecionar uma sequência distinta de comprimento r escolhida desse conjunto com repetidas seleções do mesmo elemento não sendo permitidas, **amostragem sem reposição**, é dado por

$$A_n^r = (n)_r = n(n-1)\ldots(n-r+1) = \frac{n!}{(n-r)!} = \prod_{i=0}^{r-1}(n-i),$$

desde que no primeiro procedimento (escolha do primeiro elemento da sequência) tem-se n maneiras de executá-lo, no segundo procedimento (escolha do segundo elemento da sequência) tem-se $n-1$ maneiras de executá-lo, ..., e no r-ésimo e último procedimento (escolha do r-ésimo elemento da sequência) tem-se $n-r+1$ maneiras de executá-lo. Este número de sequências é também chamado na literatura de **arranjo** quando tem-se n elementos distintos e deseja-se escolher r deles onde a ordem de escolha é importante.

2.4.2 Permutações

Um caso particular de amostragem sem reposição é quando o objetivo é saber o número de permutações de um conjunto de n elementos distintos. Neste caso, $r = n$, e o número de permutações é dado por

$$n! = (n)_n = n(n-1)\cdots 1,$$

onde $n!$ é conhecida como *função fatorial*.

Propriedades da função fatorial incluem $0! = 1! = 1$ e $n! = n(n-1)!$.

Pode-se provar (Robbins [1] (1955)) a seguinte desigualdade envolvendo a função fatorial

$$\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}e^{\frac{1}{12n}}.$$

Então, se conveniente, $n!$ pode ser aproximado por $\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}$, onde esta última fórmula é conhecida com fórmula de Stirling.

Exemplo 2.4.2: Se A é um conjunto de n elementos, quantas são as funções $f : A \to A$ injetoras?
Solução: Tem-se que garantir que cada elemento de A tem uma imagem diferente. Como A é finito e tem n elementos, f também é sobrejetora e, portanto, bijetora. Então, o primeiro elemento de A tem n opções, o segundo $n-1$ opções, até que o último elemento de A tem somente uma opção disponível. Portanto, existem $n!$ funções bijetoras $f : A \to A$. ■

Exemplo 2.4.3: De quantos modos é possível colocar r rapazes e m moças em fila de modo que as moças permaneçam juntas?
Solução: Primeiro tem-se $r+1$ opções de se escolher o lugar das moças. Em seguida, $r!$ maneiras de se escolher a posição dos rapazes entre si, e $m!$ maneiras de se escolher a posição das moças entre si. Portanto, tem-se $(r+1)r!m!$ modos diferentes de escolha. ■

Exemplo 2.4.4: Quantas são as permutaçẽs simples dos números $1, 2, \ldots, 10$ nas quais o elemento que ocupa o lugar de ordem k, da esquerda para a direita, é sempre maior que $k-3$?
Solução: Inicialmente escolhem-se os números da direita para esquerda. Observe que o número no lugar de ordem 10, tem que ser maior que 7, portanto existem 3 opções. O número no lugar de ordem 9, tem que ser maior que 6, existem, portanto, 3 opções visto que um dos núameros maiores que 6 já foi utilizado na última posição. De maneira similar pode-se ver que existem 3 opções para os números que ocupam do terceiro ao oitavo lugar. O número no lugar de ordem 2, tem somente 2 opções, pois oito números já foram escolhidos anteriormente. Finalmente, resta apenas um número para o lugar de ordem n. Portanto, existem 2×3^8 permutações deste tipo. ■

Exemplo 2.4.5: Com oito bandeiras diferentes, quantos sinais feitos com três bandeiras diferentes se podem obter?
Solução: Neste caso a ordem acarreta diferença e por isso tem-se $(8)_3 = 336$ sinais. ■

[1] H. Robbins, *A Remark of Stirling's Formula*, *Amer. Math. Monthly*, Vol. 62, No. 1, pp. 26-29, 1955.

2.4.3 Combinações

O número de coleções não ordenadas de r elementos distintos escolhidos de um conjunto universo de cardinalidade n é dado pelo *coeficiente binomial*:

$$\binom{n}{r} = \frac{(n)_r}{r!} = \frac{A_n^r}{r!} = \frac{n!}{(n-r)!r!}.$$

Para verificar isto, note que o número de coleções ordenadas de tamanho r sem repetição é $(n)_r$. Como os elementos de cada sequência de comprimento r são distintos, o número de permutações de cada sequência é $r!$. Porém, utilizando a regra da multiplicação, o procedimento de se escolher uma coleção ordenada de r termos sem repetição é igual a primeiro escolher uma coleção não-ordenada de r termos sem repetição e depois escolher uma ordem para esta coleção não-ordenada, ou seja,

$$A_n^r = (n)_r = \binom{n}{r} \cdot r!,$$

de onde segue o resultado.

O coeficiente binomial tem as seguintes propriedades:

$$\binom{n}{r} = \binom{n}{n-r},$$

$$\binom{n}{0} = 1,$$

$$\binom{n}{1} = n,$$

$$\binom{n}{r} = 0, \quad \text{se } n < r.$$

O coeficiente binomial também dá o número de subconjuntos de tamanho r que podem ser formados de um conjunto de n elementos. Como visto que o número total de subconjuntos de um conjunto de tamanho n é 2^n, então

$$2^n = \sum_{r=0}^{n} \binom{n}{r}.$$

Os números $\binom{n}{r}$ são chamados de coeficientes binomiais porque eles aparecem como coeficientes na expressão binomial $(a+b)^n$. Se n for um inteiro positivo, $(a+b)^n = (a+b)(a+b)\cdots(a+b)$. Quando a multiplicação tiver sido realizada, cada termo será formado de k elementos de a e de $(n-k)$ elementos de b, para $k = 0, 1, 2, \ldots, n$. Mas, quantos termos da forma $a^k b^{n-k}$ existirão? Simplesmente é contado o número de maneiras possíveis de escolher k dentre os n elementos a, deixando de lado a ordem (onde o i-ésimo elemento a corresponde ao i-ésimo fator do produto acima). Mas isso é justamente dado por $\binom{n}{k}$. Daí obtém-se o que é conhecido como o *Teorema Binomial*:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Exemplo 2.4.6: Dentre oito pessoas, quantas comissões de três membros podem ser escolhidas, desde que duas comissões sejam a mesma comissão se forem constituídas pelas mesmas pessoas (não se levando em conta a ordem em que sejam escolhidas)?
Solução: A resposta é dada por $\binom{8}{3} = 56$ comissões possíveis. ■

Exemplo 2.4.7: Um grupo de oito pessoas é formado de cinco homens e três mulheres. Quantas comissões de três pessoas podem ser constituídas, incluindo exatamente dois homens?
Solução: Aqui deve-se escolher dois homens (dentre cinco) e duas mulheres (dentre três). Portanto, o número procurado é $\binom{5}{2}\binom{3}{1} = 30$ comissões. ■

Exemplo 2.4.8: Quantas sequências binárias de comprimento n contém no máximo três dígitos 1?
Solução: Tem-se quatro casos possíveis: todas as sequências que não contém 1, todas as que contém apenas um 1, todas as que contém dois dígitos 1 e todas as que contém três dígitos 1. Para $0 \leq r \leq n$, existem exatamente $\binom{n}{r}$ sequências binárias com r números 1. Portanto, pela regra da adição existem

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \binom{n}{3}$$

sequências binárias de comprimento n contendo no máximo três números 1. ■

Exemplo 2.4.9: Quantas sequências de cara e coroa de comprimento n contém pelo menos 1 cara?
Solução: Neste caso, apenas uma sequência não contém qualquer cara (a sequência que contém apenas coroa). Como o número total de sequências de cara e coroa de comprimento n é igual a 2^n, então $2^n - 1$ sequências de comprimento n contém pelo menos uma cara. ■

Exemplo 2.4.10: Determine o coeficiente de x^3 no desenvolvimento de $(x^4 - \frac{1}{x})^7$.
Solução: O termo genérico do desenvolvimento é

$$\binom{7}{k}(x^4)^k(-\frac{1}{x})^{7-k} = (-1)^{7-k}\binom{7}{k}x^{5k-7}.$$

Portanto, tem-se o termo x^3 se $5k - 7 = 3$, o que implica que $k = 2$. Logo, o coeficiente de x^3 é $(-1)^5\binom{7}{2} = -21$. ■

Portanto, de modo geral, problemas de escolhas podem ser postos da seguinte forma: tem-se n objetos, entre os quais r devem ser escolhidos. Seja $\mathcal{N}$ o número de eleições possíveis. $\mathcal{N}$ depende de como a seleção é feita e do número de objetos selecionados. Segue-se a solução para cada tipo de escolha.

(a) A ordem dos elementos é relevante.

(a1) Se $r = n$ e a escolha dos objetos for sem reposição,

$$\mathcal{N} = n!,$$

isto é, tem-se a permutação dos n objetos.

(a2) Se $r < n$ e a escolha dos objetos for sem reposição,

$$\mathcal{N} = A_n^r.$$

Neste caso tem-se arranjo dos n elementos tomados r a r.

(a3) Se $r = n$, $r < n$ ou $r > n$ e a escolha dos r objetos dentre os n for com reposição,

$$\mathcal{N} = n^r.$$

(b) A ordem dos elementos é irrelevante. Neste caso, se $r < n$ e os elementos diferem na sua natureza,

$$\mathcal{N} = \binom{n}{r}.$$

2.5 Aplicações em Grafos

Grafos são modelos matemáticos utilizados para analisar a conectividade de redes. Nesta seção são descritas propriedades de grafos ilustrando uma aplicação das técnicas de contagem que foram estudadas.

2.5.1 Grafos Não Direcionados

Definição 2.5.1: Um *grafo não direcionado* $\Gamma = (V, E)$ é definido por um conjunto V de elementos chamados *nós* ou *vértices* e um conjunto $E \subseteq \{\{u, v\} : u, v \in V\}\}$ de pares não ordenados de nós que são chamados de *bordas* ou *arestas*.

Um grafo não direcionado que contém n vértices será denotado por Γ_n.

A aresta $\{u, v\}$ é vista como conectando os vértices u e v os quais são chamados de *vizinhos* ou *adjacentes*. A aresta que conecta um nó a ele mesmo é chamada de *laço*. Note que o grafo é chamado de não direcionado porque a propriedade de adjacência é simétrica[2]. Por exemplo, considere um grafo utilizado para descrever as amizades em *sites* de redes sociais. Como a relação de amizade entre as pessoas é simétrica, é necessário um grafo não direcionado para modelá-la nesses sites.

No que se segue, apenas são considerados grafos que não têm laços.

Exemplo 2.5.2: Número de grafos não direcionados com n vértices. Qual o número Γ_n de grafos não direcionados com um conjunto V de n vértices? Qual o número $\Gamma_{n,m}$ de grafos não direcionados com um conjunto V de n de vértices e um conjunto E de m arestas?

[2] Lembre-se, $=$ é simétrica porque $a = b \Rightarrow b = a$.

Solução: Note que o número de arestas é o número possível de maneiras de escolher pares de de vértices de V (a ordem dos vértices não é relevante pois o grafo é não direcionado). Então, tem-se $\binom{n}{2}$ possíveis arestas em um grafo. Cada grafo corresponde a um subconjunto do conjunto de todas as arestas. Como existem 2^r subconjuntos de um conjunto de r elementos, então existem

$$\Gamma_n = 2^{\binom{n}{2}}$$

grafos não direcionados com n vértices.

Como existem $\binom{n}{2}$ possíveis arestas, então existem

$$\Gamma_{n,m} = \binom{\binom{n}{2}}{m}$$

grafos não direcionados com n vértices e m arestas. ∎

2.5.2 Grafos Direcionados

Enquanto algumas conexões são simétricas, outras não são. Por exemplo, considere um grafo utilizado para descrever que contas do *Twitter* são seguidas umas das outras; como as relações neste caso são não simétricas, é necessário um grafo direcionado para modelá-las.

Definição 2.5.3: Um *grafo direcionado* $\Gamma = (V, E)$ é um conjunto V de vértices e um conjunto $E \subseteq \{(u, v) : u, v \in V\} = V \times V$ de pares ordenados de vértices que definem arestas direcionadas que conectam u a v, mas não necessariamente o contrário.

Exemplo 2.5.4: Quantos grafos direcionados sem laços existem com um conjunto V de n vértices? Qual o número de grafos direcionados com um conjunto V de n vértices e um conjunto E de m arestas?

Solução. Como existem $n(n-1)$ pares ordenados de vértices sem repetição, então o número total de possíveis arestas do grafo é $n(n-1)$. Cada grafo corresponde a um subconjunto do conjunto de todas as arestas. Então, existem

$$\Gamma_n = 2^{n(n-1)}$$

grafos direcionados com n vértices.

Como existem $n(n-1)$ possíveis arestas, então existem

$$\binom{n(n-1)}{m}$$

grafos direcionados com n vértices e m arestas. ∎

2.6 Contagem Multinomial ou Permutação com Elementos Repetidos

Considere r tipos de elementos e n_i cópias indistinguíveis do elemento do tipo i. Por exemplo, a palavra *probabilidade* tem duas cópias de cada uma das letras *a,b,d,i* e uma cópia de cada uma das letras *l,p,r,o,e*. O número de sequências ordenadas de comprimento $n = \sum_{i=1}^{r} n_i$ é dado por

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots 1 = \frac{n!}{\prod_{i=1}^{r} n_i!}.$$

Esta quantidade é conhecida como *coeficiente multinomial* e denotada por

$$\binom{n}{n_1\ n_2\ \ldots\ n_r},$$

onde $n = \sum_{i=1}^{r} n_i$.

Para verificar esta contagem, note que das n posições na sequência de comprimento n, pode-se escolher n_1 posições para os n_1 elementos indistinguíveis do tipo 1 de $\binom{n}{n_1}$ maneiras; das $n-n_1$ posições restantes na sequência, n_2 posições para os n_2 elementos indistinguíveis do tipo 2 de $\binom{n-n_1}{n_2}$ maneiras. Finalmente, após repetir este processo $r-1$ vezes, restam n_r posições na sequência para os n_r elementos do tipo r, que só podem ser escolhidas de uma única maneira. Utilizando o método da multiplicação, o número total de sequências possíveis é produto do número de maneiras onde os r tipos de elementos podem ser colocados.

O coeficiente multinomial também calcula o número de partições de um conjunto n elementos em r subconjuntos com tamanhos dados $n_1, n_2, \ldots, n_r$. Aplicando-se o mesmo argumento usado para demonstrar o Teorema Binomial, pode-se provar a seguinte generalização conhecida como *Teorema Multinomial*:

$$(x_1 + x_2 + \ldots + x_r)^n = \sum_{i_1=0}^{n}\sum_{i_2=0}^{n-i_1}\cdots\sum_{i_{r-1}=0}^{n-\sum_{j<r-1} i_j}\binom{n}{i_1\ i_2\ \ldots\ i_r}\prod_{k=1}^{r} x_k^{i_k},$$

onde $i_r = n - \sum_{j<r} i_j$.

Exemplo 2.6.1: Um monitor tendo resolução de $n = 1.280 \times 854$ pixels, com $r = 3$ cores possíveis (verde, azul, e vermelho) para cada pixel, pode mostrar $\binom{n}{i_1\ i_2\ i_3}$ imagens tendo i_1 pixels verdes, i_2 pixels azuis, e i_3 pixels vermelhos. O número total de imagens que pode ser exibida por este monitor para qualquer composição de cores de ver, azul, e vermelho pode ser obtido utilizando o Teorema Multinomial fazendo $x_1 = x_2 = x_3 = 1$, dando o resultado de 3^n possíveis imagens. ∎

Exemplo 2.6.2: Determine o coeficiente de x^9y^4 no desenvolvimento de $(x^3 + 2y^2 + \frac{5}{x^2})^5$.
Solução: O termo genérico do desenvolvimento é

$$\binom{5}{i_1\ i_2\ 5-i_1-i_2}(x^3)^{i_1}(2y^2)^{i_2}(\frac{5}{x^2})^{5-i_1-i_2} =$$
$$(2)^{i_2}(5)^{5-i_1-i_2}\binom{5}{i_1\ i_2\ 5-i_1-i_2}x^{3i_1-10+2i_1+2i_2}y^{2i_2}. \tag{2.1}$$

Portanto, tem-se o termo x^9y^4 se $5i_1 + 2i_2 - 10 = 9$ e $2i_2 = 4$, o que implica que $i_2 = 2$ e $i_1 = 3$. Logo, o coeficiente de x^9y^4 é $(2)^2(5)^0\binom{5}{3\ 2\ 0} = 40$. ■

2.7 Exercícios

2.1 Sabe-se que a senha pertencente a um sistema do Centro de Informática-CIn/UFPE possui 8 caracteres. Cada caracter pode ser qualquer letra (maiúsculas são diferentes de minúsculas), número ou caracter especial, somando ao todo 256 caracteres diferentes, o que corresponde aos caracteres da tabela ASC. Com base nessas informações calcule:

(a) Quantas senhas diferentes o sistema aceita?

(b) Quantas senhas diferentes podemos formar começando com a letra a?

(c) Quantas senhas diferentes contendo o número 1 podemos formar?

(d) Quantas senhas diferentes podemos ter sem repetir nenhum caracter?

(e) Quantas senhas diferentes sem caracteres repetidos possuem a letra B ou possuem o número 1 ou ambos?

(f) Desafio: Quantas senhas diferentes possuem a letra Z vindo antes o caracter $\{$? Observação: vindo antes não significa imediatamente antes. (proposto por Gustavo S. Ferreira)

2.2 O código genético especifica um aminoácido através de uma sequência de três nucleotídeos. Cada nucleotídeo pode ser de um dos quatro tipos T, A, C e G, sendo permitidas repetições. Quantos aminoácidos podem ser codificados dessa maneira?

2.3 O código Morse consiste de uma sequência de pontos e traços em que repetições são permitidas.

(a) Quantas letras se pode codificar usando exatamente n símbolos?

(b) Qual é o número de letras que se pode codificar usando n ou menos símbolos?

2.4 Um dominó é um bloco retangular dividido em dois sub-retângulos. Cada sub-retângulo possui um número. Sejam x e y esses números (não necessariamente distintos). Como o bloco é simétrico, o dominó (x, y) é igual ao dominó (y, x). Quantos blocos diferentes de dominó se pode fazer usando n números diferentes?

2.5 Um homem possui n chaves das quais, exatamente uma abre a fechadura. Ele experimenta as chaves uma de cada vez, escolhendo ao acaso em cada tentativa uma das chaves que não foi experimentada. Determine a probabilidade de que ele escolha a chave correta na r-ésima tentativa.

2.6 Uma caixa contém 40 fusíveis bons e 10 defeituosos. Suponha que se selecionam 10 fusíveis. Qual é a probabilidade de que todos eles estejam bons?

2.7 Um ônibus parte com 6 pessoas e pára em 10 pontos diferentes. Supondo que os passageiros têm igual probabilidade de descer em qualquer parada, determine a probabilidade de que dois passageiros não desembarquem na mesma parada.

2.8 Uma caixa contém 10 bolas numeradas de 1 a 10. Seleciona-se uma amostra aleatória de 3 elementos. Determine a probabilidade de que as bolas 1 e 6 estejam entre as bolas selecionadas.

2.9 Uma caixa contém b bolas pretas e r bolas vermelhas. Bolas são extraídas sem reposição, uma de cada vez. Determine a probabilidade de se obter a primeira bola preta na n-ésima extração.

2.10 Suponha que se extrai uma amostra de tamanho n de uma população de r elementos. Determine a probabilidade de que nenhum de k elementos específicos estejam na amostra se o método utilizado é

(a) amostragem sem reposição;

(b) amostragem com reposição.

2.11 Uma secretária descuidadamente coloca ao acaso n cartas em n envelopes. Determine a probabilidade de que ao menos uma carta chegue ao seu destino.

2.12 Se você possui 3 bilhetes de uma loteria para a qual se vendeu n bilhetes e existem 5 prêmios, qual é a probabilidade de você ganhar pelo menos um prêmio?

2.13 M mensagens são enviadas aleatoriamente através de N canais de comunicação, $N > M$. Encontre a probabilidade do evento

$$A = \{\text{não mais que uma mensagem seja enviada através de cada canal}\}.$$

2.14 Qual é a probabilidade de que os nascimentos de 12 pessoas caiam nos 12 diferentes meses do ano (assumindo igual probabilidade para os nascimentos nos 12 meses)?

2.15 Dez livros são colocados aleatoriamente em uma prateleira. Encontre a probabilidade de que:

(a) três particulares livros estejam sempre juntos;

(b) k particulares livros estejam sempre juntos, $2 < k < 10$.

2.16 Um conjunto de 4 *chips* de circuito integrado é constituído de 2 perfeitos e 2 defeituosos. Se 3 *chips* são selecionados aleatoriamente do grupo, qual a probabilidade do evento "dois entre os 3 selecionados são defeituosos".

2.17 Calcule a probabilidade de que algum número decimal com k dígitos escolhido aleatoriamente seja um número válido de k dígitos na base octal.

2.18 Suponha o alfabeto com 26 letras. Calcule a probabilidade de que não haja letras repetidas entre todas as seqüências com 3 letras.

2.19 Se uma caixa contém 75 *chips* de circuito integrado perfeitos e 25 defeituosos, e são selecionados aleatoriamente 12, calcule a probabilidade de que pelo menos um dentre os selecionados seja defeituoso.

2.20 Um professor faz 3 cartas de recomendação para 3 alunos. Entretanto, no momento de entregar as cartas, ao invés de entregar cada carta ao seu respectivo dono, o professor as entrega aleatoriamente.

(a) Qual é a probabilidade de que ao menos um aluno tenha recebido a carta correta?

(b) Generalize o problema para n cartas.

2.21 Em um conjunto de 5 pessoas, compute a probabilidade de que pelos menos 2 façam aniversário no mesmo dia, assumindo que o ano tem 365 dias.

2.22 De uma caixa com etiquetas numeradas de 1 a 10, retiram-se duas ao acaso, com reposição. Determine a probabilidade de que os números nas etiquetas difiram por 2.

2.23 No Brasil, a placa dos automóveis é uma *string*, na qual os 3 primeiros elementos são letras escolhidas dentre as 26, e, os 4 últimos, dígitos na base decimal.

(a) Qual é o número máximo de automóveis que podem ser emplacados neste sistema?

(b) Qual é a probabilidade de que uma placa seja iniciada pela letra K?

2.24 Uma caixa contém bolas numeradas de 1 até n.

(a) **Todas** as bolas são retiradas da caixa aleatoriamente uma a uma.

(a1) Descreva o espaço amostral.

(a2) Encontre a probabilidade de que os números selecionados sejam inteiros consecutivos em ordem crescente.

(b) Suponha a mesma caixa, com as mesmas bolas, mas agora a bola é retirada, seu número é anotado e é reposta na urna antes da retirada seguinte. Responda os itens (a1) e (a2).

2.25 M cartões de Natal são distribuídos aleatoriamente para N pessoas, $N > M$. Encontre a probabilidade de que não mais que um cartão de Natal seja enviado para cada pessoa.

2.26 Os números $1, 2, \cdots, n$ são escritos de forma aleatória. Encontre a probabilidade de que os dígitos

(a) 1 e 2,

(b) 1, 2 e 3,

apareçam como **vizinhos nessa ordem**.

(c) Repita os itens (a) e (b) considerando apenas a condição de **vizinhos**.

2.27 **(a)** Suponha que os três dígitos 1, 2 e 3 sejam escritos em ordem aleatória. Qual é a probabilidade de que ao menos um dígito ocupe seu lugar próprio?

(b) O mesmo que em (a) com os dígitos 1, 2, 3, e 4.

(c) O mesmo que em (a) com os dígitos $1, 2, \cdots, n$.

(d) Examine a resposta em (c) quando n for grande.

2.28 Suponha que de N objetos, $n < N$ sejam escolhidos ao acaso, *com* reposição. Qual será a probabilidade de que nenhum objeto seja escolhido mais do que uma vez?

2.29 Uma caixa contém etiquetas numeradas de $1, 2, \cdots, n$. Duas etiquetas são escolhidas ao acaso. Determine a probabilidade de que os números das etiquetas sejam inteiros consecutivos se:

(a) as etiquetas forem escolhidas sem reposição;

(b) as etiquetas forem escolhidas com reposição.

2.30 Dentre os números $0, 1, \cdots, 9$ são escolhidos ao acaso r números ($0 < r < 10$), com reposição. Qual é a probabilidade de que não ocorram dois números iguais?

2.31 Dois números são selecionados aleatoriamente entre os números $1, 2, \ldots, n$. Qual é a probabilidade de que a diferença entre o primeiro e o segundo números escolhidos não seja menor que m ($m > 0$).

2.32 Seja um alfabeto com 26 símbolos distintos $a, b, \cdots, z$. Considere como experimento aleatório a formação de *strings* de 3 símbolos, podendo os símbolos serem iguais.

(a) Descreva um espaço amostral para este experimento.

(b) Qual é a probabilidade de que uma *string* escolhida ao acaso dentre todas não tenha elementos repetidos?

2.33 Um lote contém n peças das quais se sabe serem r defeituosas. Se a ordem da inspeção das peças se fizer ao acaso, qual a probabilidade de que a peça inspecionada em k-ésimo lugar, $k \geq r$, seja a última peça defeituosa contida no lote?

2.34 Em um conjunto de N itens, M estão com defeito. São tomados n itens para inspeção. Se m ou mais itens dessa amostra são defeituosos, o conjunto todo é rejeitado. Encontre a probabilidade de que isto aconteça.

Capítulo 3

Probabilidade Condicional. Independência

3.1 Probabilidade Condicional

Neste capítulo tem-se duas definições de suma importância, tanto para Probabilidade e Processos Estocásticos quanto para Estatística, que são as definições de *probabilidade condicional* e *eventos independentes*. A importância e ênfase no conceito de independência ficará evidente quando você aluno descobrir como essa palavra aparecerá repetidas vezes, especialmente no contexto de variável aleatória, ou vetores aleatórios, mais para frente. Se se tem independência, problemas, de modo geral, são resolvidos facilmente. Caso contrário, são estudos de caso.

Como visto no Capítulo 1, existem várias possíveis interpretações de probabilidade. Por exemplo, pode-se interpretar probabilidade de um evento A como um limite das frequências relativas de ocorrência do evento A em realizações independentes de um experimento. Por outro lado, a interpretação subjetiva de probabilidade associa a probabilidade de um evento A com o grau de crença pessoal que o evento A ocorrerá. Em ambos os casos, probabilidade é baseada em informação e conhecimento. Revisão desta base de informação ou conhecimento pode levar a revisão do valor da probabilidade. Em particular, conhecimento que determinado evento ocorreu pode influenciar na probabilidade dos demais eventos.

Considerando-se a *interpretação frequentista* de probabilidade, suponha que o interesse seja saber qual a probabilidade do evento A, visto que sabe-se que o evento B ocorreu. Suponha que se realizasse um experimento n vezes das quais o evento A (respectivamente, B e $A \cap B$) ocorre n_A (respectivamente, $n_B > 0$ e $n_{A\cap B} \geq 0$) vezes. Seja $f_n(A) = n_A/n$ a frequência relativa do evento A nas n realizações do experimento. A probabilidade condicional de A dado que sabe-se que B ocorreu segundo esta interpretação frequentista, sugere que ela deve ser igual ao limite das frequências relativas condicionais do evento A dado o evento B, isto é, deve ser o limite da razão $n_{A\cap B}/n_B$ quando n tende ao infinito. É fácil provar que esta razão é igual a $f_n(A \cap B)/f_n(B)$, que por sua vez segundo a interpretação frequentista de probabilidade é aproximadamente igual a $P(A \cap B)/P(B)$ para valores grandes de n.

Considerando-se uma *interpretação subjetiva*, suponha que a incerteza de um agente é descrita por uma probabilidade P em $(\Omega, \mathcal{A})$ e que o agente observa ou fica sabendo que o evento B ocorreu. Como o agente deve atualizar sua probabilidade $P(\cdot|B)$ de modo a incorporar esta nova informação? Claramente, se o agente *acredita* que B é verdadeiro, então parece razoável requerer que

$$P(B^c|B) = 0. \tag{3.1}$$

Em relação aos eventos contidos em B, é razoável assumir que sua chance relativa permaneça inalterada se tudo que o agente descobriu foi que o evento B ocorreu, ou seja, se $A_1, A_2 \subseteq B$ com $P(A_2) > 0$, então

$$\frac{P(A_1)}{P(A_2)} = \frac{P(A_1|B)}{P(A_2|B)}. \tag{3.2}$$

Segue que (3.1) e (3.2) determinam completamente $P(\cdot|B)$ se $P(B) > 0$.

Teorema 3.1.1: *Se $P(B > 0)$ e $P(\cdot|B)$ é uma medida de probabilidade em Ω que satisfaz (3.1) e (3.2), então*

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

Prova: Como $P(\cdot|B)$ é uma medida de probabilidade e satisfaz $P(B^c|B) = 0$, então $P(B|B) = 1 - P(B^c|B) = 1$. Considerando $A_1 = A$ e $A_2 = B$ em (3.2), logo $P(A|B) = \frac{P(A)}{P(B)}$ para $A \subseteq B$. Se A não é um subconjunto de B, tem-se que $A = (A \cap B) \cup (A \cap B^c)$. Como $(A \cap B)$ e $(A \cap B^c)$ são eventos disjuntos, $P(A|B) = P(A \cap B|B) + P(A \cap B^c|B)$. Como $A \cap B^c \subseteq B^c$ e $P(B^c|B) = 0$, então $P(A \cap B^c|B) = 0$. Como $A \cap B \subseteq B$, usando o caso anterior

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

∎

Deste modo as interpretações frequentista e subjetivista de probabilidade justificam a seguinte definição.

Definição 3.1.2: Seja $(\Omega, \mathcal{A}, P)$ um espaço de probabilidade. Se $A, B \in \mathcal{A}$ e $P(B) > 0$ a *probabilidade condicional de A dado B* é definida por

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

Para um evento fixo B que satisfaz $P(B) > 0$, $P(\cdot|B)$ satisfaz aos axiomas K1-K4 (Capítulo 1) e realmente é uma medida de probabilidade. Para provar $K2$, note que para todo $A \in \mathcal{A}$, como $P(A \cap B) \geq 0$,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0.$$

Para provar $K3$, como $\Omega \cap B = B$, então

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Finalmente, para provar (K5)$'$ (que implica K4), se $A_1, A_2, \ldots$ são mutuamente exclusivos $A_1 \cap B, A_2 \cap B, \ldots$ também o são, então

$$\begin{aligned} P(\cup_i A_i|B) &= \frac{P((\cup_i A_i) \cap B)}{P(B)} \\ &= \frac{P(\cup_i (A_i \cap B))}{P(B)} \\ &= \frac{\sum_i P(A_i \cap B)}{P(B)} \\ &= \sum_i P(A_i|B). \end{aligned}$$

Logo, para um dado evento B fixo, tal que $P(B) > 0$, tem-se que $P(\cdot|B)$ satisfaz todas as propriedades de uma probabilidade incondicional vistas no Capítulo 1.

A probabilidade condicional também satisfaz às seguintes propriedades facilmente prováveis:

(i) $P(B|B) = 1$.

(ii) $P(A|B) = P(A \cap B|B)$.

(iii) Se $A \supseteq B$, então $P(A|B) = 1$.

(iv) $P(A \cap B|C) = P(A|B \cap C)P(B|C)$.

(v) Fazendo $C = \Omega$ na propriedade (iv) acima,

$$P(A \cap B) = P(A|B)P(B).$$

Utilizando indução matemática, pode-se facilmente provar que

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)\ldots P(A_n|A_1 \cap \ldots \cap A_{n-1}).$$

Exemplo 3.1.3: Suponha uma turma com 30 alunos dos quais 5 são mulheres e o restante homens. 4 alunos saem da sala de aula sucessivamente. Qual a probabilidade de que sejam 2 mulheres e 2 homens, nessa ordem?

Solução: Sejam os eventos:

$$M_1 = \{\text{o primeiro a sair é mulher}\},$$

$$M_2 = \{\text{o segundo a sair é mulher}\},$$

$$H_3 = \{\text{o terceiro a sair é homem}\},$$

$$H_4 = \{\text{o quarto a sair é homem}\}.$$

Portanto, a probabilidade pedida é:

$$P(M_1 \cap M_2 \cap H_3 \cap H_4) = P(M_1)P(M_2|M_1)\ldots P(H_4|M_1 \cap M_2 \cap H_3) = \frac{5}{30}\frac{4}{29}\frac{25}{28}\frac{24}{27} = \frac{100}{5481}.$$

■

Um método de se obter uma probabilidade (incondicional) de uma probabilidade condicional é utilizando o *Teorema da Probabilidade Total.*

Teorema 3.1.4: *Seja a sequência de eventos $B_1, B_2, \ldots$ uma partição de Ω. então para todo $A \in \mathcal{A}$*

$$P(A) = \sum_{i:P(B_i)\neq 0} P(A|B_i)P(B_i).$$

Prova:

Como $B_1, B_2, \ldots$ é uma partição de Ω,

$$A = A \cap \Omega = A \cap (\cup_i B_i) = \cup_i (A \cap B_i).$$

Como os eventos B_i's são mutuamente exclusivos, os eventos $(A \cap B_i)$'s também são mutuamente exclusivos. Então o axioma (K5)$'$ implica que

$$\begin{aligned} P(A) &= P(\cup_i (A \cap B_i)) \\ &= \sum_i P(A \cap B_i) \\ &= \sum_{i:P(B_i)\neq 0} P(A \cap B_i) \\ &= \sum_{i:P(B_i)\neq 0} P(A|B_i)P(B_i). \end{aligned}$$

■

Se os eventos da partição $B_1, B_2, \ldots$ são interpretados como possíveis causas e o evento A corresponda a um efeito particular associado a uma causa, $P(A|B_i)$ especifica a relação estocástica entre a causa B_i e o efeito A.

Exemplo 3.1.5: Seja $\{D, D^c\}$ uma partição do espaço amostral, onde o evento D significa que um dado indivíduo possui uma certa doença. Seja A o evento que determinado teste para o diagnóstico da doença deu positivo. Então, $P(A|D^c)$ descreve a probabilidade do exame dá positivo mesmo que o paciente esteja saudável, é a chamada probabilidade de *falso positivo*. $P(A^c|D)$ é a probabilidade do exame dá negativo mesmo que o paciente esteja doente, é a chamada probabilidade de *falso negativo*. Estas probabilidades determinam a qualidade do teste, quanto menores as probabilidades de falso negativo e falso positivo melhor a qualidade do teste. Caso as probabilidades $P(D), P(A|D), P(A|D^c)$ sejam conhecidas pode-se usando o Teorema da Probabilidade Total obter a probabilidade incondicional de determinado exame dar positivo $P(A)$. ■

Porém, geralmente o que se busca é saber que dado que o resultado de um exame deu positivo qual a probabilidade de que o indivíduo esteja doente. Pode-se obter esta probabilidade utilizando a famosa *fórmula de Bayes*:

$$P(D|A) = \frac{P(A \cap D)}{P(A \cap D) + P(A \cap D^c)} = \frac{P(A|D)P(D)}{P(A|D)P(D) + P(A|D^c)P(D^c)}.$$

Mais geralmente, a **fórmula de Bayes** é dada por:

$$\begin{aligned} P(B_i|A) &= \frac{P(A \cap B_i)}{\sum_j P(A \cap B_j)} \\ &= \frac{P(A \cap B_i)}{\sum_{j:P(B_j)\neq 0} P(A \cap B_j)} \\ &= \frac{P(A|B_i)P(B_i)}{\sum_{j:P(B_j)\neq 0} P(A|B_j)P(B_j)}. \end{aligned}$$

Os B_i podem descrever, por exemplo, diferentes mensagens emitidas em um sistema de comunicações e A pode descrever uma mensagem recebida pelo sistema. $P(A|B_i)$ determina a probabilidade que a mensagem B_i seja emitida e a mensagem A seja recebida por este sistema. Essas probabilidades condicionais especificam o modelo do canal de comunicações. Caso as probabilidades $P(B_i)$'s de cada mensagem ser enviada e as probabilidades condicionais que descrevem o canal de comunicação sejam conhecidas pode-se usando o Teorema da Probabilidade Total obter a probabilidade incondicional que determinada mensagem A seja recebida. Entretanto, de modo geral, o objetivo é saber que, se uma dada mensagem foi recebida (efeito), A, qual a probabilidade de cada uma das mensagens B_i terem sido as mensagens enviadas. Podem-se obter estas probabilidades utilizando-se a *fórmula de Bayes*.

É fácil de provar a fórmula de Bayes usando o Teorema da Probabilidade Total. As probabilidades $P(B_i)$ são usualmente chamadas de probabilidades *a priori* e as probabilidades condicionais $P(B_i|A)$ de probabilidades *a posteriori*. O seguinte exemplo, extraído de Fine (2006), ilustra uma aplicação da fórmula de Bayes.

Exemplo 3.1.6: Considere uma imagem formada por $n \times m$ pixels com a k-ésima linha contendo d_k ($\leq m$) pixels defeituosos. No primeiro estágio do experimento uma linha é escolhida ao acaso. A seguir, um pixel é selecionado ao acaso nessa linha e constatado ser defectivo; seja D este evento. Qual a probabilidade de que este pixel defeituoso esteja na linha k?

Solução: Seja $R = k$ o evento que este pixel pertencia a k-ésima linha da imagem. A fórmula de Bayes permite determinar que, dado que

$$P(R = k) = \frac{1}{n}$$

e

$$P(D|R=k)=\frac{d_k}{m},$$

tem-se que

$$P(R=k|D)=\frac{\frac{1}{n}\frac{d_k}{m}}{\sum_{i=1}^{n}\frac{1}{n}\frac{d_i}{m}}=\frac{d_k}{\sum_{i=1}^{n}d_i}.$$

■

Exemplo 3.1.7: Um sistema de comunicação telegráfico transmite os sinais ponto (.) e traço (-). A experiência tem mostrado que 2/5 dos pontos e 1/3 dos traços são mudados. Suponha que a razão entre os pontos transmitidos e os traços transmitidos é de 5 para 3. Qual é a probabilidade de que o sinal recebido seja o que foi transmitido quando

(a) o sinal recebido é um ponto;

(b) o sinal recebido é um traço.

Solução:
Sejam os eventos

$$R_\bullet=\{\text{um ponto é recebido}\},$$

$$R_\_=\{\text{um traço é recebido}\},$$

$$T_\bullet=\{\text{um ponto é transmitido}\},$$

$$T_\_=\{\text{um traço é transmitido}\}.$$

e as probabilidades dadas no problema ou decorrentes de usar o complementar: $P(R_\bullet\mid T_\bullet)=\frac{3}{5}$, $P(R_\bullet\mid T_\_)=\frac{1}{3}$, $P(R_\_\mid T_\bullet)=\frac{2}{5}$, $P(R_\_\mid T_\_)=\frac{2}{3}$, $P(T_\bullet)=\frac{5}{8}$ e $P(T_\_)=\frac{3}{8}$.

Tem-se que:

$$R_\bullet=(R_\bullet\cap T_\bullet)\cup(R_\bullet\cap T_\_),$$

$$R_\_=(R_\_\cap T_\_)\cup(R_\_\cap T_\bullet),$$

logo,

$$P(R_\bullet)=P(R_\bullet\mid T_\bullet)P(T_\bullet)+P(R_\bullet\mid T_\_)P(T_\_)=\frac{3}{5}\frac{5}{8}+\frac{1}{3}\frac{3}{8}=\frac{4}{8},$$

$$P(R_\_)=P(R_\_\mid T_\_)P(T_\_)+P(R_\_\mid T_\bullet)P(T_\bullet)=\frac{2}{3}\frac{3}{8}+\frac{2}{5}\frac{5}{8}=\frac{4}{8}.$$

(a)

$$P(T_\bullet\mid R_\bullet)=\frac{P(R_\bullet\cap T_\bullet)}{P(R_\bullet)}=\frac{3}{4}.$$

(b)

$$P(T\_ \mid R\_) = \frac{P(T\_ \cap R\_)}{P(R\_)} = \frac{1}{2}.$$

■

Exemplo 3.1.8: Um canal de comunicação binário envia um dentre dois tipos de sinais, denotados por 0 e 1. Devido ao ruído, um 0 transmitido é alguma vezes recebido como um 1 e um 1 transmitido é alguma vezes recebido como um 0. Para um dado canal, assuma uma probabilidade de 0.94 que um 0 transmitido seja corretamente recebido como um 0 e uma probabilidade de 0.91 que um 1 transmitido seja corretamente recebido como um 1. Adicionalmente, assuma uma probabilidade de 0.45 de se transmitir um 0. Se um sinal é enviado, determine,

(a) A probabilidade de que um 1 seja recebido.

(b) A probabilidade de que um 0 seja recebido.

(c) A probabilidade de que um 1 foi transmitido, dado que um 1 foi recebido.

(d) A probabilidade de que um 0 foi transmitido, dado que um zero foi recebido.

(e) A probabilidade de um erro.

Solução:
Sejam os eventos

$$T_0 = \{\text{um 0 é transmitido}\},$$

$$T_1 = \{\text{um 1 é transmitido}\},$$

$$R_0 = \{\text{um 0 é recebido}\},$$

$$R_1 = \{\text{um 1 é recebido}\}.$$

Logo,

$$P(R_0 \mid T_0) = 0.94 \Rightarrow P(R_1 \mid T_0) = 0.06,$$

$$P(R_1 \mid T_1) = 0.91 \Rightarrow P(R_0 \mid T_1) = 0.09,$$

$$P(T_0) = 0.45,$$

$$P(T_1) = 0.55.$$

(a)

$$R_1 = (R_1 \cap T_1) \cup (R_1 \cap T_0),$$

logo,

$$P(R_1) = P(R_1 \mid T_1)P(T_1) + P(R_1 \mid T_0)P(T_0) = 0.91 \times 0.55 + 0.06 \times 0.45 = 0.5275.$$

(b)

$$R_0 = (R_0 \cap T_0) \cup (R_0 \cap T_1),$$

logo,

$$P(R_0) = P(R_0 \mid T_0)P(T_0) + P(R_0 \mid T_1)P(T_1) = 0.94 \times 0.45 + 0.09 \times 0.55 = 0.4725,$$

ou,

$$P(R_0) = 1 - P(R_1) = 1 - 0.5275 = 0.4725.$$

(c)

$$\begin{aligned} P(T_1 \mid R_1) &= \frac{P(T_1 \cap R_1)}{P(R_1)} \\ &= \frac{P(R_1 \mid T_1)P(T_1)}{P(R_1)} \\ &= \frac{0.91 \times 0.55}{0.5275} = 0.9488. \end{aligned}$$

(d)

$$\begin{aligned} P(T_0 \mid R_0) &= \frac{P(T_0 \cap R_0)}{P(R_0)} \\ &= \frac{P(R_0 \mid T_0)P(T_0)}{P(R_0)} \\ &= \frac{0.94 \times 0.45}{0.4725} = 0.8952. \end{aligned}$$

(e)

$$E = \{\text{acontece um erro}\}.$$

Logo,

$$E = (T_1 \cap R_0) \cup (T_0 \cap R_1),$$

$$P(E) = P(R_0 \mid T_1)P(T_1) + P(R_1 \mid T_0)P(T_0) = 0.09 \times 0.55 + 0.06 \times 0.45 = 0.0765.$$

■

Exemplo 3.1.9: Uma urna contém 4 bolas brancas e 6 bolas pretas. Sacam-se, sucessivamente e sem reposição, duas bolas dessa urna. Determine a probabilidade da primeira bola ser branca sabendo que a segunda bola é branca.

Solução: Sejam B_1 e B_2 os eventos a primeira bola é branca e a segunda bola é branca, respectivamente. Queremos calcular $P(B_1|B_2)$. Utilizando a fórmula de Bayes,

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

Mas $P(B_2|B_1) = \frac{3}{9}$, $P(B_2|B_1^c) = \frac{4}{9}$, $P(B_1) = \frac{4}{10}$ e $P(B_1^c) = \frac{6}{10}$. Logo,

$$P(B_1|B_2) = \frac{\frac{3}{9} \cdot \frac{4}{10}}{\frac{3}{9} \cdot \frac{4}{10} + \frac{4}{9} \cdot \frac{6}{10}} = \frac{\frac{2}{15}}{\frac{2}{5}} = \frac{1}{3}.$$

■

Exemplo 3.1.10: Se $P(C|D) = 0,4$ e $P(D|C) = 0,5$, que evento é mais provável C ou D?
Solução:

$$P(C \mid D) = \frac{P(C \cap D)}{P(D)} = 0.4 \Rightarrow P(D) = \frac{P(C \cap D)}{0.4}.$$

$$P(D \mid C) = \frac{P(C \cap D)}{P(C)} = 0.5 \Rightarrow P(C) = \frac{P(C \cap D)}{0.5}.$$

Como $\frac{P(C\cap D)}{0.4} > \frac{P(C\cap D)}{0.5}$, então D é mais provável que C. ■

Exemplo 3.1.11: Se $P(E) = 0.4$ e $P(F) = 0.7$, o que pode-se concluir sobre $P(E|F)$?
Solução: Por definição,

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Porém,

$$\max\{P(E) + P(F) - 1, 0\} \leq P(E \cap F) \leq \min\{P(E), P(F)\}.$$

Logo, $0.1 \leq P(E \cap F) \leq 0.4$, portanto

$$\frac{0.1}{0.7} \leq P(E|F) \leq \frac{0.4}{0.7}.$$

■

Os dois próximos exemplos foram extraídos de Fine (2005) e servem para ilustrar aplicações de probabilidade condicional que desafiam nossa intuição.

Exemplo 3.1.12: **(Paradoxo de Monty Hall)** Monty Hall foi um popular apresentador de programa de jogos em TV cujo jogo começava mostrando ao participante três portas fechadas d_1, d_2, d_3, onde atrás de apenas uma delas havia um prêmio valioso. O participante selecionava uma porta, por exemplo, d_1, mas antes que a porta fosse aberta, Monty Hall, que sabia em que porta estava o prêmio, por exemplo, d_2, abria a porta restante d_3, que não continha o prêmio. O participante tinha então permissão para ficar com sua porta original, d_1, ou escolher a outra porta fechada. A pergunta é se é melhor ficar com a porta original ou trocar de porta.
Solução: A fórmula de Bayes é utilizada para analisar este problema. Seja G uma porta escolhida aleatoriamente para conter o prêmio; Y a porta que o participante escolhe primeiro; e M a porta que Monty Hall abre. O participante não tem qualquer conhecimento a priori sobre a localização do prêmio, ou seja ele considera todas as portas equiprováveis, e isto pode ser modelado por

$$P(G = d_i|Y = d_j) = \frac{1}{3},$$

isto é, todas as portas têm a mesma probabilidade de conter o prêmio não importa qual porta o participante escolha. Se o participante escolher uma porta que não contém o prêmio, Monty Hall necessariamente terá de abrir a porta que não contém o prêmio, isto pode ser modelado por

$$P(M = d_{i_1}|Y = d_{i_2}, G = d_{i_3}) = 1,$$

onde $i_1, i_2, i_3 \in \{1, 2, 3\}$ e são distintos. Se o participante escolher corretamente, por exemplo, $Y = G = d_{i_2}$, então Monty Hall escolhe aleatoriamente entre as outras duas outras portas:

$$P(M = d_{i_1}|Y = G = d_{i_2}) = \frac{1}{2}, \text{ para } d_{i_1} \neq d_{i_2}.$$ [1]

Para determinar se o participante deve trocar de porta, deve-se calcular

$$\begin{aligned}
P(G = d_1|Y = d_2, M = d_3) &= \frac{P(G = d_1, Y = d_2, M = d_3)}{P(Y = d_2, M = d_3)} \\
&= \frac{P(M = d_3|G = d_1, Y = d_2)P(G = d_1|Y = d_2)P(Y = d_2)}{P(M = d_3|Y = d_2)P(Y = d_2)} \\
&= \frac{P(M = d_3|G = d_1, Y = d_2)P(G = d_1|Y = d_2)}{P(M = d_3|Y = d_2)} \\
&= \frac{1/3}{P(M = d_3|Y = d_2)}.
\end{aligned}$$

O Teorema da Probabilidade Total e a definição de probabilidade condicional são utilizados para determinar o valor de $P(M = d_3|Y = d_2)$.

$$\begin{aligned}
&P(M = d_3|Y = d_2) = \frac{P(Y = d_2, M = d_3)}{P(Y = d_2)} \\
&= \frac{P(Y = d_2, M = d_3, G = d_1) + P(Y = d_2, M = d_3, G = d_2) + P(Y = d_2, M = d_3, G = d_3)}{P(Y = d_2)} \\
&= \frac{P(M = d_3|Y = d_2, G = d_1)P(G = d_1|Y = d_2)P(Y = d_2)}{P(Y = d_2)} \\
&+ \frac{P(M = d_3|Y = d_2, G = d_2)P(G = d_2|Y = d_2)P(Y = d_2)}{P(Y = d_2)} \\
&+ \frac{P(M = d_3|Y = d_2, G = d_3)P(G = d_3|Y = d_2)P(Y = d_2)}{P(Y = d_2)} \\
&= P(M = d_3|Y = d_2, G = d_1)P(G = d_1|Y = d_2) \\
&+ P(M = d_3|Y = d_2, G = d_2)P(G = d_2|Y = d_2) \\
&+ P(M = d_3|Y = d_2, G = d_3)P(G = d_3|Y = d_2) \\
&= 1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 = \frac{1}{2}.
\end{aligned}$$

[1] A solução depende como este caso é resolvido.

Logo, $P(G = d_1|Y = d_2, M = d_3) = \frac{2}{3}$, e o participante deve trocar de porta de sua escolha original d_2 para d_1! ■

Exemplo 3.1.13: Seja D o evento que um indivíduo selecionado ao acaso de uma população tem uma doença particular. A probabilidade que um indivíduo selecionado ao acaso nesta população tenha determinada doença é p_d. Existe um teste para diagnóstico desta doença que sempre acusa presença da doença quando o indivíduo tem a doença. Contudo, quando o indivíduo não tem a doença, o teste reporta falsamente que o indivíduo tem a doença com probabilidade p_t. Seja TP o evento que o teste reporta positivamente que o indivíduo tem a doença. Formalmente,

$$P(D) = p_d, P(TP|D) = 1, P(TP|D^c) = p_t.$$

Um indivíduo pode estar interessado em saber a probabilidade $P(D|TP)$ que ele tenha a doença dado que o teste deu positivo. Se, por exemplo, a doença for rara, $p_d = 0,001$ e o teste reportar falsamente com probabilidade pequena $p_t = 0,05$, será visto que, apesar desta pequena probabilidade do teste dá um resultado errado, a probabilidade do indivíduo ter a doença é pequena.

Solução:

Pela fórmula de Bayes

$$P(D|TP) = \frac{P(TP|D)P(D)}{P(TP|D)P(D) + P(TP|D^c)P(D^c)} = \frac{p_d}{p_d + p_t(1 - p_d)} = 0,02.$$

■

Exemplo 3.1.14: Suponha que todos os bytes tenham a mesma probabilidade de ocorrência. Seja W o número de 1's em um byte. Considere os seguintes eventos:

$$A = \{\text{O primeiro e o segundo bit são iguais a 1}\}$$

e

$$B = \{W \text{ é um número ímpar}\}.$$

Calcular $P(A)$, $P(B)$, $P(B|A)$ e $P(A|B)$.

Solução:

$$P(A) = \frac{||A||}{||\Omega||} = \frac{2^6}{2^8} = \frac{1}{4}.$$

$$P(B) = \frac{||B||}{||\Omega||} = \frac{\binom{8}{1} + \binom{8}{3} + \binom{8}{5} + \binom{8}{7}}{2^8} = \frac{1}{2}.$$

$$P(B|A) = \frac{P(A \cap B}{P(A)},$$

onde $P(A \cap B) = \frac{||A \cap B||}{\Omega} = \frac{\binom{6}{1}+\binom{6}{3}+\binom{6}{5}}{2^8} = \frac{1}{8}$. Portanto,

$$P(B|A) = \frac{\frac{1}{8}}{\frac{1}{4}} = \frac{1}{2}.$$

$$P(A|B) = \frac{P(A \cap B)}{B} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}.$$

■

Exemplo 3.1.15: Dois dados são jogados, um após o outro, e observa-se o evento a soma dos dois dados é igual a 9; então, qual a probabilidade do primeiro dado ter dado resultado 4?

Solução:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{36}}{\frac{4}{36}} = \frac{1}{4}.$$

■

De acordo com Halpern (2003), embora probabilidade condicional seja bastante útil, ela sofre de problemas, em particular quando se quer tratar de eventos de probabilidade zero. Tradicionalmente, se $P(B) = 0$, então $P(A|B)$ não é definida. Isto leva a um número de dificuldades filosóficas em relação a eventos com probabilidade zero. São eles realmente impossíveis? Caso contrário, quão improvável um evento precisa ser antes de ele ser atribuído probabilidade zero? Deve um evento em algum caso ser atribuído probabilidade zero? Se existem eventos com probabilidade zero que não são realmente impossíveis, então o que significa condicionar em eventos de probabilidade zero? Por exemplo, considere o espaço de probabilidade $([0,1], \mathcal{B}, \mu)$ onde $\mathcal{B}$ é a σ-ágebra de Borel restrita a eventos contidos em $[0,1]$ e μ é uma medida de probabilidade na qual todo intervalo em $[0,1]$ possui probabilidade igual ao seu comprimento. Seja $B = \{1/4, 3/4\}$ e $A = \{1/4\}$. Como $P(B) = 0$, $P(A|B)$ não é definida. Porém parece razoável assumir que neste caso $P(A|B) = 1/2$ já que μ intuitivamente implica que todos os estados são equiprováveis, mas a definição formal de probabilidade condicional não permite obter esta conclusão. Para ver uma alternativa à teoria descrita neste capítulo ver Halpern (2003).

3.2 Independência

Intuitivamente, dois eventos são independentes se eles nada têm a ver um com o outro, são não relacionados; a ocorrência de um não tem qualquer influência sobre a ocorrência do outro. A intuição por trás da frase "o evento A é independente do evento B" é que o conhecimento sobre a tendência para A ocorrer dado que sabe-se que B ocorreu não é alterada quando sabe-se que B ocorreu. Então, usando probabilidades condicionais pode-se formalizar esta intuição da seguinte forma: A é independente de B se $P(A|B) = P(A)$. Mas usando a definição de probabilidade condicional,

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

Como esta última expressão é definida inclusive para o caso de $P(B) = 0$, ela é a expressão adotada como a definição de independência entre dois eventos.

Definição 3.2.1: O evento A é independente do evento B se $P(A \cap B) = P(A)P(B)$.

Esta definição de independência implica que independência é um conceito simétrico em teoria da probabilidade, isto é, A é independente de B se e somente se B é independente de A. Note que esta definição também implica que eventos A e B são independentes se $P(A) = 0$ ou $P(B) = 0$, o que pode gerar conclusões não intuitivas se de fato $P(A) = 0$ ou $P(B) = 0$. Por exemplo, se $P(A) = 0$, então A é independente dele mesmo, porém A certamente não é não relacionado consigo mesmo. Similarmente, é fácil provar que se $P(A) = 1$, A é independente dele mesmo. O seguinte teorema prova que estes são os únicos casos em que um evento é independente dele mesmo.

Teorema 3.2.2: *A é independente dele mesmo se e somente se $P(A) = 0$ ou $P(A) = 1$.*

Prova:
$$P(A \cap A) = P(A) = P(A)P(A) \Leftrightarrow P(A) = 0 \text{ ou } P(A) = 1.$$

∎

Intuitivamente, se A é independente de B o fato que B não ocorreu, ou seja que B^c ocorreu, não deve alterar a probabilidade de A. Portanto, é de se esperar que se A e B são independentes, então A e B^c também são. O seguinte teorema prova que esta intuição é verdadeira.

Teorema 3.2.3: *Se A e B são eventos independentes, A e B^c, respectivamente A^c e B e A^c e B^c também o são.*

Prova:
$$A = A \cap \Omega = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c).$$

Então, como $A \cap B$ e $A \cap B^c$ são mutuamente exclusivos, o axioma K3 implica que

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

Como A e B são independentes,

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

Rearrajando os termos e utilizando o fato que $P(B^c) = 1 - P(B)$, tem-se que $P(A \cap B^c) = P(A)P(B^c)$. ∎

O conceito de independência também se aplica a uma coleção arbitrária de eventos $\{A_i\}_{i \in \mathcal{I}}$, onde $\mathcal{I}$ é um conjunto de índices. Neste caso, têm-se duas definições.

Definição 3.2.4: Uma coleção de eventos $\{A_i\}_{i \in \mathcal{I}}$ é *independente par a par* se para todo $i \neq j \in I$, A_i e A_j são eventos independentes.

Definição 3.2.5: Uma sequência finita de eventos $A_1, A_2, \ldots, A_n$, $n \geq 1$, é *mutuamente independente* se para todo $I \subseteq \{1, \ldots, n\}$,

$$P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i).$$

Definição 3.2.6: Uma coleção de eventos $\{A_i\}_{i \in \mathcal{I}}$ é *mutuamente independente* se para todo $J \subseteq \mathcal{I}$ finito, $\{A_i\}_{i \in J}$ são mutuamente independentes.

O seguinte exemplo, extraído de James (1996), ilustra que independência par a par não implica independência mútua.

Exemplo 3.2.7: Se $\Omega = \{1, 2, 3, 4\}$ e $P(\{w\}) = 1/4$, então $A = \{1, 2\}$, $B = \{1, 3\}$, e $C = \{2, 3\}$ são eventos independentes par a par.
Solução: Pode-se verificar isto pelo fato que

$$P(A \cap B) = P(\{1\}) = \frac{1}{4} = \frac{1}{2}\frac{1}{2} = P(A)P(B).$$

Similarmente, pode-se provar o mesmo resultado para os outros pares. Contudo,

$$P(A \cap B \cap C) = P(\emptyset) = 0 \neq P(A)P(B)P(C) = \frac{1}{8}.$$

Então, A, B, e C não são mutuamente independentes. ■

Exemplo 3.2.8: Um sistema automático de alarme contra incêndio utiliza três células sensíveis ao calor, que agem, independentemente, uma da outra. Cada célula entra em funcionamente com probabilidade 0.8 quando a temperatura atinge 60^oC. Se pelo menos uma das células entrar em funcionamente, o alarme soa. Calcule a probabilidade do alarme soar.
Solução: Sejam

$$C_i = \{\text{célula } i \text{ entra em funcionamento}, i = 1, \cdots, 3.\}$$

Logo,

$$P(C_1 \cup C_2 \cup C_3) = 1 - P(\overline{C_1 \cup C_2 \cup C_3}) = 1 - P(\overline{C_1})P(\overline{C_2})P(\overline{C_3}) = 1 - 0.2^3 = 0.992.$$

■

Exemplo 3.2.9: Se $\Omega = \{1, 2, 3, 4, 5, 6\}$, $A = \{1, 2, 4\}$, e $B = \{2, 3, 5\}$, então construa uma medida de probabilidade em Ω tal que A e B sejam independentes.
Solução: Seja p_i a probabilidade do elemento $i \in \Omega$. Então, para que A e B sejam independentes,

$$P(A \cap B) = p_2 = P(A)P(B) = (p_1 + p_2 + p_4)(p_2 + p_3 + p_5).$$

Por exemplo, pode-se escolher $p_1 = p_2 = p_3 = p_6 = \frac{1}{4}$ e $p_4 = p_5 = 0$. Deste modo, $P(A \cap B) = \frac{1}{4}$ e $P(A) = P(B) = \frac{1}{2}$. ■

Exemplo 3.2.10: O evento F de que um determinado sistema falhe ocorre se os eventos A_1 ou A_2 ocorrerem, mas o evento A_3 não ocorrer. Se A_1, A_2, A_3 são mutumente independentes e $P(A_1) = 0.4$, $P(A_2) = 0.35$, e $P(A_3) = 0.1$, então calcule $P(F)$.
Solução: O evento F é igual ao evento $(A_1 \cup A_2) \cap A_3^c$. Logo sua probabilidade é igual a:

$$P(F) = P((A_1 \cup A_2) \cap A_3^c) = P(A_1 \cup A_2)P(A_3^c)$$
$$= (P(A_1) + P(A_2) - P(A_1)P(A_2))(1 - P(A_3)) = (0.4 + 0.35 - 0,4 \cdot 0.35)(0.9) = 0.549.$$

■

Exemplo 3.2.11: Assuma que $A_1, \ldots, A_n$ são eventos mutuamente independentes e que $P(A_i) = p_i$. Calcular as probabilidades dos seguintes eventos:

(a) O evento A é o evento onde todos estes eventos ocorrem:

$$P(A) = P(\cap_{i=1}^n A_i) = \prod_{i=1}^n P(A_i) = \prod_{i=1}^n p_i.$$

(b) O evento B é o evento que nenhum desses eventos ocorre:

$$P(B) = P(\cap_{i=1}^n A_i^c) = \prod_{i=1}^n P(A_i^c) = \prod_{i=1}^n (1 - p_i).$$

(c) O evento C é o evento onde pelo menos um desses eventos ocorre:

$$P(C) = P(B^c) = 1 - P(B) = 1 - \prod_{i=1}^n (1 - p_i).$$

■

Exemplo 3.2.12: João e José disputam um jogo com uma moeda equilibrada. Cada jogador lança a moeda duas vezes e vence o jogo aquele que primeiro obtiver dois resultados iguais. João começa jogando e se não vencer passa a moeda para José e continuam alternando jogadas. Qual a probabilidade de João vencer o Jogo?
Solução: Seja A_k o evento dois resultados iguais são obtidos na k-ésima tentativa. Note que $P(A_k) = \frac{1}{2}$. Seja B_k o evento João ganha na sua k-ésima jogada. Então,

$$B_1 = A_1;\; B_2 = A_1^c \cap A_2^c \cap A_3;\; B_3 = A_1^c \cap A_2^c \cap A_3^c \cap A_4^c \cap A_5,$$

em geral,

$$B_k = A_1^c \cap A_2^c \cap \cdots \cap A_{2k-2}^c \cap A_{2k-1}.$$

Portanto,

$$P(B_k) = P(A_1^c \cap A_2^c \cap \cdots \cap A_{2k-2}^c \cap A_{2k-1}) = P(A_1^c)P(A_2^c) \cdots P(A_{2k-2}^c)P(A_{2k-1}) = (\frac{1}{2})^{2k-1},$$

onde a penúltima igualdade se deve ao fato dos lançamentos serem independentes. Logo,

$$P(\text{João vencer}) = P(\cup_{k=1}^{\infty} B_k) = \sum_{k=1}^{\infty} P(B_k) = \sum_{k=1}^{\infty} (\frac{1}{2})^{2k-1} = \frac{2}{3}.$$

■

3.3 Exercícios

3.1 Sabe-se que os eventos B_1, B_2 e B_3 são disjuntos par a par e que sua união é igual ao espaço amostral. Estes eventos têm as probabilidades $P(B_1) = 0.2$ e $P(B_2) = 0.3$. Existe um outro evento A tal que $P(A|B_1) = 0.3$, $P(A|B_2) = 0.4$ e $P(A|B_3) = 0.1$. Calcule:

(a) $P(A)$.

(b) $P(B_2|A)$.

3.2 Considere os eventos A, B e C. Sendo A e B independentes, A e C independentes e B e C mutuamente excludentes, mostre que A e $B \cup C$ são independentes.

3.3 Sejam $A_1, A_2, \ldots A_n$ eventos independentes com $p_k = P(A_k)$, $k = 1, \ldots, n$. Obtenha a probabilidade de ocorrência dos seguintes eventos, em termos das probabilidades p_k:

(a) A ocorrência de nenhum dos A_k.

(b) A ocorrência de pelo menos um dos A_k.

(c) A ocorrência de exatamente um dos A_k.

(d) A ocorrência de exatamente dois dos A_k.

(e) A ocorrência de, no máximo, $n-1$ dos A_k.

3.4 Numa certa cidade, 75% de seus habitantes têm menos de 30 anos, enquanto os outros 25% têm mais de 30 anos. Sabendo-se que a taxa de alfabetização entre os jovens, idade < 30 anos é de 40% e entre os não jovens, idade ≥ 30 anos, é de 30%, calcule:

(a) a probabilidade de que um habitante escolhido ao acaso seja alfabetizado;

(b) a probabilidade de que um habitante alfabetizado ter menos de 30 anos.

3.5 Um centro de processamento de dados comprou um lote de 5000 *chips*, dos quais 1000 foram manufaturados pela fábrica A e o restante pela B. Sabe-se que 10% dos *chips* produzidos por A e 5% dos produzidos por B, respectivamente, são defeituosos.

(a) Um *chip* é escolhido aleatoriamente do lote. Qual é a probabilidade de que seja defeituoso?

(b) Um *chip* é escolhido aleatoriamente do lote, observado, e constata-se que é defeituoso. Qual é a probabilidade de que tenha sido produzido por A?

(c) Suponha que uma amostra de 20 *chips* seja retirada aleatoriamente do lote comprado. Qual será a probabilidade de se encontrar na amostra pelo menos 1 defeituoso? (este item será facilmente resolvido usando uma Binomial, a qual será vista posteriormente)

3.6 Um porta-níqueis contém moedas de prata e de cobre em igual número. Extraem-se ao acaso e sem reposição duas moedas. Calcule a probabilidade de que:

(a) saia uma moeda de prata na segunda tiragem;

(b) uma e uma só das moedas seja de prata;

(c) a segunda moeda extraída seja de prata, sabendo-se que a primeira era de cobre;

(d) pelo menos uma das moedas seja de cobre.

3.7 Seja o espaço amostral $\Omega = \{a, b, c, d, e\}$ onde $P(\{a, b, c\}) = \frac{1}{2}$ e $P(\{a\}) = \frac{1}{4}$.

(a) Determine as probabilidades de todos os eventos cujas probabilidades podem ser computadas dos dados.

(b) Compute $P(\{b, c, d\} \mid \{a, b, c\})$.

(c) Compute $P(\{a\} \mid \{a, b, c\})$.

3.8 Sabe-se que em um centro de processamento de dados, 80% dos programas são escritos em C, 20% em Haskell, e que 20% dos programas em C e 40% dos em Haskell compilam da primeira vez.

(a) Qual é a probabilidade de que um programa selecionado aleatoriamente compile da primeira vez?

(b) Se um programa selecionado aleatoriamente compilar da primeira vez, qual é a probabilidade de que tenha sido escrito em Haskell?

3.9 Suponha que a ocorrência ou não de chuva dependa das condições do tempo no dia imediatamente anterior. Admita que se chove hoje, choverá amanhã com probabilidade 0.7 e que se não chove hoje choverá amanhã com probabilidade 0.4. Sabendo-se que choveu hoje, calcule a probabilidade que choverá depois de amanhã.

3.10 Em um teste de múltipla escolha, a probabilidade do aluno saber a resposta é p. Havendo m escolhas se ele sabe a resposta, responde corretamente com probabilidade 1; se não sabe, responde corretamente com probabilidade $1/m$.

(a) Qual a probabilidade de que a pergunta tenha sido respondida corretamente?

(b) Qual a probabilidade que o aluno sabia a resposta dado que a pergunta foi respondida corretamente?

3.11 Sejam $A_1, A_2, \ldots A_n$ eventos independentes com $p_k = P(A_k)$, $k = 1, \ldots, n$. Obtenha a probabilidade de ocorrência dos seguintes eventos, em termos das probabilidades p_k:

(a) A ocorrência de nenhum dos A_k.

(b) A ocorrência de pelo menos um dos A_k.

3.12 Considere as seis permutações das letras a, b, c como também os *triplets* (a, a, a), (b, b, b), (c, c, c). Seja Ω consistindo dos nove *triplets*, cada um com probabilidade $1/9$. Definindo os eventos

$$A_k = \{ \text{ o } k\text{-ésimo lugar é ocupado pela letra } a \ \},$$

para $k = 1, \cdots, 3$ mostre que eles são independentes dois a dois mas não são independentes três a três (a questão também poderia ter sido: verifique se os eventos são mutuamente independentes).

3.13 Suponha que três rapazes possuem bonés idênticos. Cada um atira seu boné no centro de uma mesa. Os bonés são misturados e então cada um seleciona aleatoriamente um boné.

(a) Qual é a probabilidade que nenhum dos três tenha escolhido seu próprio boné?

(b) Resolva o mesmo problema para n.

3.14 Durante um dado período de tempo, um radar detecta um alvo com probabilidade p. Sabe-se que as detecções de alvos por períodos de tempo idênticos, são independentes umas das outras. Encontre a probabilidade que o míssel seja detectado em ao menos um dos n períodos de tempo idênticos.

3.15 Um computador consiste de n unidades. A confiabilidade (tempo livre de falha) da 1a. unidade durante o tempo T é p_1, da 2a. unidade para o tempo T é p_2, e assim por diante. As unidades falham independentemente umas das outras. Quando qualquer unidade falha, o computador falha. Encontre a probabilidade de que o computador falhe durante o tempo T.

3.16 Três mensagens são enviadas através de três canais de comunicação, cada uma das quais pode ser transmitida com diferente exatidão. A transmissão de uma mensagem pode levar a um dos seguintes eventos:

$$A_1 = \{\text{ a mensagem é transmitida da forma correta}\};$$
$$A_2 = \{\text{ a mensagem é parcialmente distorcida}\};$$
$$A_3 = \{\text{ a mensagem é completamente distorcida}\}.$$

As probabilidades dos eventos A_1, A_2 e A_3 são conhecidas e iguais a p_1, p_2 e p_3 $(p_1 + p_2 + p_3 = 1)$. Considerando que mensagens podem ser distorcidas ou transmitidas corretamente independentemente umas das outras, encontre a probabilidade dos seguintes eventos:

$$A = \{\text{todas as três mensagens são transmitidas da forma correta}\}.$$

$$B = \{\text{pelo menos uma das mensagens é completamente distorcida}\}.$$

3.17 Durante um dado período de tempo, um *software* pode apresentar erros com probabilidade p_0. Assumindo independência entre os eventos considerados, quantos períodos de tempo são necessários para que erros sejam detectados com probabilidade não menor que p?

3.18 Uma mensagem que está sendo transmitida através de um canal de comunicação consiste de n símbolos. Durante a transmisssão, a probabilidade de cada um dos símbolos serem distorcidos independentemente uns dos outros, é p. Por questões de segurança, cada mensagem é então enviada k vezes.

(a) Encontre a probabilidade de que pelo menos uma das mensagens que está sendo transmitida, não seja distorcida em qualquer um dos seus símbolos.

(b) Quantas vezes uma mensagem precisa ser repetida para que a probabilidade de que pelo menos uma das mensagens não seja distorcida não seja menor que p?

3.19 Coloque V ou F nas sentenças abaixo:

(a) A e B independentes $\Rightarrow P(A \cup B) = P(A) + P(B)$. ( )

(b) A e B independentes $\Rightarrow P(A \cup B) = P(A) + P(B) - P(A)P(B)$. ( )

(c) A e B independentes $\Rightarrow P(A \cap B) = P(A)P(B)$. ( )

(d) A e B independentes $\Rightarrow P(A \mid B) = P(B)$. ( )

(e) A e B independentes $\Rightarrow P(A \mid B) = P(A)$. ( )

(f) A e B são excludentes $\Leftrightarrow$ A e B são independentes. ( )

Nos itens a seguir $B = \{\text{bebo}\}$, $D = \{\text{dirijo}\}$. Você vai responder estes itens tendo em vista que você é um **cidadão brasileiro responsável, consciente de que o futuro do seu país depende de você**, aliás, você **é o futuro do Brasil**!

(g) $B \cap \overline{D}$. ( )

(h) $\overline{B} \cap D$. ( )

(i) $P(D \mid B) = 1$. ( )

(j) $P(B \mid D) = 0$. ( )

(k) B e D são eventos independentes. ( )

(l) B e D são eventos excludentes. ( )

3.20 Uma mensagem consistindo de n símbolos binários 0 e 1 é enviada. Cada símbolo é distorcido com uma probabilidade p. Por questões de segurança a mensagem é repetida duas vezes. A informação é considerada correta se ambas as mensagens coincidem. Encontre a probabilidade de que ambas as mensagens estejam distorcidas, a despeito de coincidirem.

3.21 A causa de um acidente está sendo investigada e existem quatro hipótesis possíveis: H_1, H_2, H_3 e H_4. Estatisticamente sabe-se que $P(H_1) = 0.2$, $P(H_2) = 0.4$, $P(H_3) = 0.3$ e $P(H_4) = 0.1$. Já é sabido que ocorreu o evento $A = \{\text{falha no nível do óleo}\}$. Pelas mesmas estatísticas a probabilidade condicional do evento A dadas as hipótesis H_1, H_2, H_3 e H_4 são, respectivamente, 0.9, 0, 0.2 e 0.3. Encontre as probabilidades a posteriori para as hipótesis.

3.22 Um colégio é composto de 70% de homens e 30% de mulheres. Sabe-se que 40% dos homens e 60% das mulheres são fumantes. Qual é a probabilidade de que um estudante que foi visto fumando seja homem? (estes dados, atualmente, pelo menos entre os alunos do CCEN e do CIn, ambos da UFPE são irreais, pois as probabilidades de fumantes são quase zero!)

3.23 Suponha que os automóveis têm igual probabilidade de serem produzidos na segunda, terça, quarta, quinta e sexta-feira. As percentagens de automóveis amarelos produzidos nos diferentes dias da semana são: segunda, 4%; terça, quarta e quinta, 1%; sexta, 2%. Se você compra um automóvel amarelo, qual é a probabilidade de que o mesmo foi produzido numa segunda-feira?

3.24 Um homem dispara 12 tiros independentemente num alvo. Qual é a probabilidade de que atinja o alvo pelo menos uma vez, se tem probabilidade 9/10 de atingir o alvo em qualquer tiro?

3.25 Certo experimento consiste em lançar um dado equilibrado duas vezes, independentemente. Dado que os dois números sejam diferentes, qual é a probabilidade (condicional) de:

(a) pelo menos um dos números ser 6;

(b) a soma dos números ser 8.

3.26 Três prisioneiros[2] são informados por seu carcereiro que um deles foi escolhido aleatoriamente para ser executado, e os outros dois serão libertados. O prisioneiro A pede ao carcereiro para lhe dizer confidencialmente qual, de seus dois companheiros de cela, será libertado, afirmando que não há qualquer problema, pois ele ja sabe que pelo menos um deles estará em liberdade. O carcereiro recusa-se a responder a pergunta, argumentando que, se A soubesse qual de seus companheiros seria libertado, então sua própria probabilidade de ser executado cresceria de 1/3 para 1/2. Que você pensa do julgamento de carcereiro?

3.27 Num *stand* de automóveis os registros indicam que 50% dos clientes pretendem ar condicionado no carro, 49% preferem carro com direção hidráulica e 25% interessam-se pelas duas coisas simultaneamente. Um registro é selecionado aleatoriamente.

(a) Qual é a probabilidade de que o ar condicionado tenha sido pretendido mas não a preferência do carro com direção hidráulica?

(b) Qual é a probabilidade de que nenhuma das referidas preferências tenha sido selecionada?

(c) Qual é a probabilidade de exatamente uma das referidas preferências ter sido selecionada?

3.28 Três jornais A, B e C são publicados em uma cidade e uma recente pesquisa entre os eleitores indica o seguinte: 20% lêem A; 26% lêem B; 14% lêem C; 8% lêem A e B; 5% lêem A e C; 2% lêem A, B e C; 4% lêem B e C. Para um adulto escolhido ao acaso, calcule a probabilidade de que:

[2]Este problema aparece em vários livros dentre os quais: S. M. Ross, *Introduction to Probability Models.* Fifth Edition, Academic Press, 1972; R. Isaac, *The Pleasures of Probability.* Springer-Verlag, 1995; F. Mosteller, *Fifty Challenging Problems in Probability.* Dover Publications, Inc., New York, 1965; S. Russel and P. Norvig, *Artifitial Intelligence A Modern Approach.* Prentice Hall, New Jersey, 1995. Você vê alguma semelhança entre o citado problema e o Paradoxo de Monty Hall?

(a) ele não leia qualquer dos jornais;

(b) ele leia exatamente um dos jornais;

(c) ele leia ao menos A e B se se souber que ele lê ao menos um dos jornais publicados.

3.29 Uma máquina impressora pode imprimir n letras, digamos $\alpha_1, \alpha_2, \cdots \alpha_n$. Ela é acionada por impulsos elétricos, cada letra sendo produzida por um impulso diferente. Suponha que exista uma probabilidade constante p de imprimir a letra correta e também suponha independência. Um dos n impulsos, escolhido ao acaso, foi alimentado na máquina duas vezes e, em ambas, a letra α_1 foi impressa. Calcule a probabilidade de que o impulso escolhido tenha sido para imprimir α_1.

3.30 Estima-se que a probabilidade de que Mário seja culpado é 0.2. São chamadas duas testemunhas, Alberto e Carlos. Se Mário for realmente culpado, Alberto dirá que ele é culpado com certeza e Carlos dirá que Mário é culpado com probabilidade 0.6. Se Mário for inocente, Alberto dirá com probabilidade de 0.3 que ele é inocente e Carlos dirá certamente que ele é inocente.

(a) Qual é a probabilidade de Alberto dizer que Mário é inocente?

(b) Qual é a probabilidade de Mário ser inocente se Carlos disser que é inocente?

3.31 Em um centro de computação científica os *software* NET_1.0, NET_2.0 e NET_3.0 são igualmente responsáveis por todas as simulações realizadas. Um pesquisador tem observado que, das simulações provenientes de cada *software*, 8, 6 e 4 porcento, respectivamente, apresentam erros. Ele escolhe ao acaso uma simulação.

(a) Qual é a probabilidade de que contenha erros?

(b) Qual é a probabilidade de que não seja de NET_1.0, sabendo-se que tem erro?

Capítulo 4

Variáveis Aleatórias Unidimensionais e Funções

4.1 Introdução

Analisando o tráfego de redes Ethernet, o interesse pode ser, por exemplo, nas variáveis número total de *bytes*, ou número total de pacotes, ou ainda, percentual de utilização da rede em determinados períodos de tempo. A Tabela 4.1 exibe variáveis de interesse em uma rede de computadores.

Tabela 4.1: Variáveis aleatórias observadas na camada de aplicação de redes de computadores.

Aplicação	Variável observada
Vídeoconferência	Número de *frames* por segundo
	Número de pixels por *frame*
	Intervalo de tempo entre chegadas de pacotes ou *bytes* ou *bits*
	Número de pacotes ou de *bytes* ou de *bits* por intervalo de tempo
WWW	Intervalo de tempo entre sessões WWW
	Tamanho do documento (arquivo) WWW
	Número de imagens embutidas em um documento
	Intervalo de tempo entre chegadas de pedidos WWW
	Localidade temporal do *site*
	Popularidade do *site*
	Intervalo de tempo entre chegadas de pacotes ou *bytes* ou *bits*
	Número de pacotes ou de *bytes* ou de *bits* por intervalo de tempo
Correio eletrônico	Intervalo de tempo entre sessões de correio eletrônico
	Tamanho da mensagem de e-mail
	Intervalo de tempo entre chegadas de pacotes ou *bytes* ou *bits*
	Número de pacotes ou de *bytes* ou de *bits* por intervalo de tempo

Suponha que um experimento aleatório consista em selecionar ao acaso um perfil em uma rede social. Quantos amigos na rede social possui o dono deste perfil? Quantidades desse tipo é o que tradicionalmente têm sido chamadas de *variáveis aleatórias*. Intuitivamente, são variáveis aleatórias porque seus valores variam dependendo do perfil escolhido. O adjetivo "aleatória" é usado para enfatizar que o seu valor é de certo modo incerto.

Formalmente, como será visto a seguir, uma variável aleatória não é nem variável, nem aleatória. Na verdade, variáveis aleatórias são funções, reais. Uma variável aleatória é uma função real. Sequências de variáveis aleatórias são sequências de funções reais. Convergência de variáveis aleatórias é convergência de funções reais e teoremas limite sobre variáveis aleatórias são teoremas limite sobre funções reais.

Definição 4.1.1: Seja $(\Omega, \mathcal{A}, P)$ um espaço de probabilidade. Uma função real $X : \Omega \to R$, é chamada de *variável aleatória* se para todo Boreliano B, $X^{-1}(B) \in \mathcal{A}$, onde $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ é o conjunto de elementos do espaço amostral cuja imagem segundo X está em B.

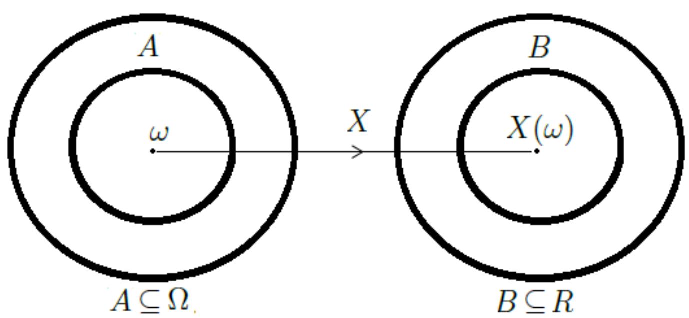

Figura 4.1: Esquema funcional de uma variável aleatória.

Notações comumente encontradas, com os respectivos significados:

$$[X = x] = \{\omega \in \Omega \mid X(\omega) = x\},\ B = \{x\},$$

$$[X \leq x] = \{\omega \in \Omega \mid X(\omega) \leq x\},\ B = (-\infty, x],$$

$$[x \leq X \leq y] = \{\omega \in \Omega \mid x \leq X(\omega) \leq y\},\ B = [x, y].$$

4.2 Definindo uma probabilidade sobre $\mathcal{B}$

Dada uma variável aleatória X, pode-se definir uma probabilidade, P_X, no espaço mensurável $(\mathbb{R}, \mathcal{B})$ da seguinte maneira: para todo $B \in \mathcal{B}$, seja $P_X(B) = P(X^{-1}(B))$. Por definição de variável aleatória, tem-se que $X^{-1}(B) \in \mathcal{A}$, então P_X está bem definida. P_X satisfaz os axiomas K2, K3, e K5$'$ de probabilidade, pois:

(K2) $P_X(B) = P(X^{-1}(B)) = P(A) \geq 0.$

(K3) $P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1.$

(K5′) Suponha que $B_1, B_2, \ldots$ são eventos Borelianos disjuntos dois a dois. Então,

$$P_X(\cup_n B_n) = P(X^{-1}(\cup_n B_n)) = P(\cup_n(X^{-1}(B_n))) = \sum_n P(X^{-1}(B_n)) = \sum_n P_X(B_n).$$

A probabilidade P_X é dita como sendo a *probabilidade induzida* pela variável aleatória X e um outro espaço de probabilidade é $(\mathbb{R}, \mathcal{B}, P_X)$.

4.3 Função de Distribuição Acumulada

Para uma dada variável aleatória X, uma maneira de descrever a probabilidade induzida P_X é utilizando sua *função de distribuição acumulada.*

Definição 4.3.1: A *função de distribuição acumulada* de uma variável aleatória X, representada por F_X, é definida por

$$F_X(x) = P(X \leq x) = P_X((-\infty, x]), \forall x \in \mathbb{R}.$$

A função de distribuição acumulada F_X satisfaz às seguintes propriedades:

(F1) Monotonicidade. Se $x \leq y$, então $F_X(x) \leq F_X(y)$.

$$x \leq y \Rightarrow (-\infty, x] \subseteq (-\infty, y] \Rightarrow P_X((-\infty, x]) \leq P_X((-\infty, y]) \Rightarrow F_X(x) \leq F_X(y).$$

(F2) Continuidade à direita. Se $x_n \downarrow x$, então $F_X(x_n) \downarrow F_X(x)$.

Se $x_n \downarrow x$, então os eventos $(-\infty, x_n]$ são não-crescentes e $\cap_n(-\infty, x_n] = (-\infty, x]$. Logo, pela continuidade da probabilidade, tem-se que $P_X((-\infty, x_n]) \downarrow P((-\infty, x])$, ou seja, $F_X(x_n) \downarrow F_X(x)$.

(F3) Se $x_n \downarrow -\infty$, então $F_X(x_n) \downarrow 0$, e se $x_n \uparrow \infty$, então $F_X(x_n) \uparrow 1$.

Se $x_n \downarrow -\infty$, então os eventos $(-\infty, x_n]$ são não-crescentes e $\cap_n(-\infty, x_n] = \emptyset$. Logo, pela continuidade da probabilidade, tem-se que $P_X((-\infty, x_n]) \downarrow P(\emptyset)$, ou seja, $F_X(x_n) \downarrow 0$. Similarmente, se $x_n \uparrow \infty$, então os eventos $(-\infty, x_n]$ são não-decrescentes e $\cup_n(-\infty, x_n] = \mathbb{R}$. Logo, pela continuidade da probabilidade, tem-se que $P_X((-\infty, x_n]) \uparrow P(\Omega)$, ou seja, $F_X(x_n) \uparrow 1$.

Teorema 4.3.2: *Uma função real F satisfaz F1–F3 se e somente se F é uma função de distribuição acumulada.*

Prova: A prova de que se F for uma função de distribuição acumulada, então F satisfaz F1-F3 foi dada acima. A prova de que toda função real que satisfaz F1-F3 é uma função acumulada está fora do escopo deste livro (Burrill, 1982)(Shryayev, 1984). ∎

Uma função de distribuição acumulada pode corresponder a várias variáveis aleatórias no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$. Por exemplo, seja X tal que $P(X = 1) = P(X = -1) = \frac{1}{2}$. Logo, $P(-X = 1) = P(-X = -1) = \frac{1}{2}$. Portanto, X e $-X$ têm a mesma distribuição. Consequentemente, $F_X = F_{-X}$.

Teorema 4.3.3: *Seja D o conjunto de pontos de descontinuidade da função de distribuição F_X. Então, D é enumerável.*

Prova: Pela monotonicidade, tem-se que para todo $x \in \mathbb{R}$, $F_X(x^-) \leq F_X(x) \leq F_X(x^+)$. Logo, $x \in D$ se, e somente se, $F_X(x^+) > F_X(x^-)$. Para $n = 1, 2, 3, \ldots$ seja

$$A_n = \{x : F_X(x^+) - F_X(x^-) > \frac{1}{n}\}.$$

Então, $D = \cup_{n=1}^{\infty} A_n$. Será visto que todo A_n contém menos que n pontos e, portanto, é finito, dessa forma, D seré enumerável.

Por absurdo, suponha que exista A_n contendo n pontos. Assim, $A_n = \{x_1, x_2, \ldots, x_n\}$, onde $x_1 < x_2 < \cdots x_n$ e

$$0 \leq F_X(x_1^-) \leq F_X(x_1^+) \leq F_X(x_2^-) \leq F_X(x_2^+) \leq \cdots \leq F_X(x_n^-) \leq F_X(x_n^+) \leq 1.$$

Então, $\sum_{k=1}^{n}(F_X(x_k^+) - F_X(x_k^-)) \leq 1$. Mas por definição do conjunto A_n, tem-se que $F_X(x_i^+) - F_X(x_i^-) > \frac{1}{n}$ para todo $x_i \in A_n$. Portanto, $\sum_{k=1}^{n}(F_X(x_k^+) - F_X(x_k^-)) > n \times \frac{1}{n} > 1$, absurdo. Logo, A_n contém menos que n pontos. ∎

Exemplo 4.3.4: Este exemplo mostra como usar a função de distribuição acumulada para calcular probabilidades. Lembrando que

$$F_X(x) = P(X \leq x) = P_X((-\infty, x]).$$

(a)

$$\begin{gathered}(-\infty, b] = (-\infty, a] \cup (a, b],\ a \leq b \Rightarrow \\ P((-\infty, b]) = P((-\infty, a]) + P((a, b]) \Rightarrow \\ P((a, b]) = P((-\infty, b]) - P((-\infty, a]) = F_X(b) - F_X(a) \Rightarrow\end{gathered}$$

$$P_X((a, b]) = P(a < X \leq b) = F_X(b) - F_X(a). \tag{4.1}$$

(b) $I_n = \{x : a - \frac{1}{n} < x \leq a + \frac{1}{n}\}$. Isto significa que $I_1 \supset I_2 \supset \cdots \Rightarrow \lim_{n\to\infty} I_n = \cap_{n=1}^{\infty} I_n = \{a\}$. Pela continuidade monotônica da probabilidade[1] tem-se que

$$\begin{aligned} P(X = a) &= P(\cap_{n=1}^{\infty} I_n) \\ &= P(\lim_{n\to\infty} I_n) \\ &= \lim_{n\to\infty} P(I_n) \\ &= \lim_{n\to\infty} P(a - \frac{1}{n} < X \leq a + \frac{1}{n}) \\ &= \lim_{n\to\infty} (F_X(a + \frac{1}{n}) - (F_X(a - \frac{1}{n})) \\ &= \lim_{n\to\infty} F_X(a + \frac{1}{n}) - \lim_{n\to\infty} F_X(a - \frac{1}{n}) \Rightarrow \end{aligned}$$

[1]ou lembrando do Capítulo 1 que $P(\lim I_n) = \lim P(I_n)$,

$$P_X(a) = P(X = a) = F_X(a^+) - F_X(a^-) = F_X(a) - F_X(a^-). \tag{4.2}$$

A expressão (4.2) é o salto da função de distribuição no ponto a. Se X é uma variável aleatória discreta assumindo o valor a, $F_X(a) - F_X(a^-) \geq 0$.

(c)

$$(a, b) \cup \{b\} = (a, b] \Rightarrow$$
$$P((a, b)) + P(\{b\}) = P((a, b]) \Rightarrow$$
$$P((a, b)) = P((a, b]) - P(b) = F_X(b) - F_X(a) - P(X = b) \Rightarrow$$

$$P_X((a, b)) = P(a < X < b) = F_X(b) - F_X(a) - P(X = b). \tag{4.3}$$

O resultado em (4.3) foi obtido usando (4.1) e (4.2).

(d)

$$(a, b] \cup \{a\} = [a, b] \Rightarrow$$
$$P((a, b]) + P(a) = P([a, b]) \Rightarrow$$
$$P([a, b]) = P((a, b]) + P(X = a) = F_X(b) - F_X(a) + P(X = a) \Rightarrow$$

$$P_X([a, b]) = P(a \leq X \leq b) = F_X(b) - F_X(a) + P(X = a). \tag{4.4}$$

O resultado em (4.4) foi obtido usando (4.1) e (4.2).

(e)

$$[a, b) = (a, b) \cup \{a\} \Rightarrow$$
$$P([a, b)) = P((a, b)) + P(a) = F_X(b) - F_X(a) - P(X = b) + P(X = a) \Rightarrow$$

$$P_X([a, b)) = P(a \leq X < b) = F_X(b) - F_X(a) - (P(X = b) + P(X = a)). \tag{4.5}$$

(4.5) foi obtida a partir de ((4.3)).

(f)

$$(-\infty, b] = (-\infty, b) \cup \{b\} \Rightarrow$$
$$P((-\infty, b]) = P((-\infty, b)) + P(X = b) \Rightarrow$$
$$P((-\infty, b)) = P((-\infty, b]) - P(X = b) \Rightarrow$$

$$P_X(-\infty, b)) = P(-\infty < X < b) = F_X(b) - P(X = b). \tag{4.6}$$

■

4.4 Tipos de Variáveis Aleatórias

Existem três tipos de variáveis aleatórias: **discreta**, **contínua** e **singular**.

4.4.1 Variável Aleatória Discreta

Definição 4.4.1: Uma variável aleatória X é *discreta* se assume valores num conjunto enumerável com probabilidade 1, ou seja, se existe um conjunto enumerável $\{x_1, x_2, \ldots\} \subseteq \mathbb{R}$ tal que $P(X = x_i) \geq 0,\ \forall i \geq 1$ e $P(X \in \{x_1, x_2, \ldots\}) = 1$.

A função $p(\cdot)$ definida por

$$p(x_i) = P_X(\{x_i\}), i = 1, 2, \ldots$$

e

$$p(x) = 0, x \notin \{x_1, x_2, \ldots\},$$

é chamada de *função probabilidade* (de massa) de X. Toda função probabilidade é uma função real e assume valores entre 0 e 1, sendo positiva para uma quantidade enumerável de pontos e tal que $\sum_i p(x_i) = 1$. De modo geral escreve-se

$$0 \leq p(x_i) \leq 1,$$

$$\sum_i p(x_i) = 1.$$

O conjunto de pontos

$$(x_i, p(x_i)), i = 1, 2, \ldots,$$

é usualmente denotado na literatura por *distribuição de probabilidade* da variável aleatória X.

Para esta variável aleatória tem-se que

$$F_X(x) = \sum_{i: x_i \leq x} p(x_i),$$

de modo que F_X é uma função degrau. Seja $p : \mathbb{R} \to [0, 1]$, sendo p positiva para uma quantidade enumerável de pontos $\{x_1, x_2, \ldots\}$ e satisfazendo $\sum_i p(x_i) = 1$ e seja

$$P(B) = \sum_{x_i \in B} p(x_i), \forall B \in \mathcal{B}.$$

Prova-se que $P(B)$ é uma probabilidade em $(R, \mathcal{B})$ (P satisfaz os axiomas de Kolmogorov). Logo, a distribuição de uma variável aleatória discreta X pode ser determinada tanto pela função de distribuição acumulada F_X quanto pela sua função de probabilidade p.

Exemplo 4.4.2: Seja X uma variável aleatória assumindo os valores $1, 2, \ldots$, com $P(X = k) = \frac{1}{2^k}$, $k = 1, 2, \ldots,$. Portanto, a distribuição de probabilidade da variável aleatória X é

$$(k, \frac{1}{2^k}), k = 1, 2, \cdots,$$

■

Exemplo 4.4.3: Este exemplo mostra como calcular a função de distribuição acumulada para uma variável aleatória discreta. Seja X assumindo os valores $0, 1, 2$ com igual probabilidade. Portanto,

$$x < 0 \Rightarrow F_X(x) = 0,$$

$$0 \leq x < 1 \Rightarrow F_X(x) = P(X = 0) = \frac{1}{3},$$

$$1 \leq x < 2 \Rightarrow F_X(x) = P(X = 0) + P(X = 1) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3},$$

$$x \geq 2 \Rightarrow F_X(x) = P(X = 0) + P(X = 1) + P(X = 2) = 1.$$

Assim,

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{3}, & 0 \leq x < 1, \\ \frac{2}{3}, & 1 \leq x < 2, \\ 1, & x \geq 2. \end{cases}$$

■

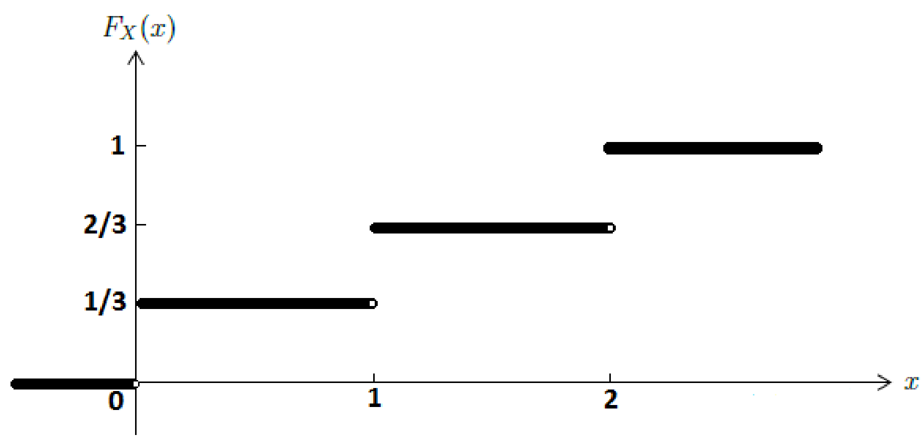

Figura 4.2: Gráfico de função de distribuição acumulada para uma variável aleatória discreta

Exemplo 4.4.4: Este exemplo mostra como calcular a função probabilidade de X a partir do conhecimento da função de distribuição acumulada. Seja

$$F_X(x) = \begin{cases} 0, & x < -1, \\ \frac{1}{6}, & -1 \leq x < 1, \\ \frac{2}{3}, & 1 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$$

logo,

$$P(X = -1) = F_X(-1) - F_X(-1^-) = \frac{1}{6} - 0 = \frac{1}{6},$$

$$P(X = 1) = F_X(1) - F_X(1^-) = \frac{2}{3} - \frac{1}{6} = \frac{1}{2},$$

$$P(X = 3) = F_X(3) - F_X(3^-) = 1 - \frac{2}{3} = \frac{1}{3}$$

$$P(X = k) = 0, \forall k \neq -1, 1, 3.$$

■

4.4.2 Variável Aleatória Contínua

Definição 4.4.5: Uma variável aleatória X é *contínua* se existe uma função real $f_X(x) \geq 0$ tal que

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \forall x \in R.$$

A função f_X é chamada de *função densidade de probabilidade* de X. F_X é contínua e

$$f_X(x) = F'_X(x).$$

Uma função $f(x) \geq 0$ é densidade de alguma variável aleatória se e somente se, $\int_{-\infty}^{\infty} f(x)dx = 1$, sendo neste caso fácil provar que a função F definida por $\int_{-\infty}^{x} f(t)dt$ satisfaz às condições F1, F2, e F3. Portanto, pelo Teorema 4.3.2, F é uma função de distribuição acumulada. Então, para variável aleatória contínua, a distribuição de uma variável aleatória contínua X pode ser determinada tanto pela função de distribuição acumulada F_X quanto pela sua função densidade f_X.

Uma variável aleatória X tem densidade se F_X é a integral (de Lebesgue) de sua derivada; sendo, neste caso, a derivada de F_X uma função densidade para X. Em quase todos os casos encontrados na prática, uma variável aleatória X tem densidade se F_X é (i) contínua e (ii) derivável por partes, ou seja, se F_X é derivável no interior de um número finito ou enumerável de intervalos cuja união é $\mathbb{R}$.

Por exemplo, seja

$$F_X(x) = \begin{cases} 0, & \text{se } x < 0, \\ x^2, & \text{se } 0 \leq x < 1, \\ 1, & \text{se } x \geq 1. \end{cases}$$

Então X tem densidade pois F_X é contínua e derivável em todos os pontos da reta exceto em $\{0, 1\}$.

Quando X é uma variável aleatória contínua,

$$\begin{aligned} P(X < b) &= F_X(b) - P(X = b) \\ &= F_X(b) - (F_X(b) - F_X(b^-)) \\ &= F_X(b) \\ &= P(X \leq b). \end{aligned}$$

Exemplo 4.4.6: Seja X uma variável aleatória com densidade

$$f_X(x) = \begin{cases} 2(m^2+1)x + (k+1)m + (k+1)^2, & 0 < x < 1, \\ 0, & x \notin (0,1), \end{cases}$$

onde m e k são números reais.

(a) Determine os valores de m e k especificando então $f_X(x)$.

(b) Obtenha $F_X(x)$.

Solução:

(a)

$$\int_0^1 (2(m^2+1)x + (k+1)m + (k+1)^2)dx = 1 \Rightarrow$$

$$m^2 + m(k+1) + (k+1)^2 = 0 \Rightarrow$$

$$m = \frac{-(k+1) \pm \sqrt{(k+1)^2 - 4(k+1)^2}}{2}.$$

Como

$$k \in \mathbb{R} \Rightarrow -3(k+1)^2 \geq 0 \Rightarrow k = -1.$$

Logo, $m = 0$. Portanto,

$$f_X(x) = \begin{cases} 2x, & 0 < x < 1, \\ 0, & x \notin (0,1), \end{cases}$$

(b)

$$x < 0, F_X(x) = 0,$$

$$0 \leq x < 1, F_X(x) = \int_{-\infty}^{x} 2tdt = x^2,$$

$$x > 1, F_X(x) = 1.$$

Logo,

$$F_X(x) = \begin{cases} 0, & x < 0, \\ x^2, & x \in [0,1), \\ 1, & x \geq 1. \end{cases}$$

■

4.4.3 Variável Aleatória Singular

Definição 4.4.7: Uma variável aleatória X é *singular* se F_X é uma função contínua cujos pontos de crescimento formam um conjunto de comprimento (medida de Lebesgue) nulo.

Na Seção 4.7 é apresentado um exemplo de variável aleatória singular.

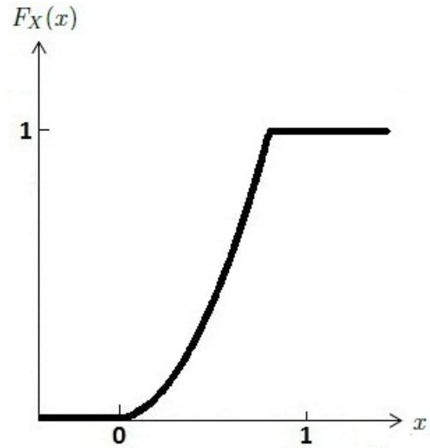

Figura 4.3: Gráfico de função de distribuição acumulada para uma variável aleatória contínua

4.4.4 Variável Aleatória Mista

Na prática, a maioria das variáveis aleatórias é discreta ou contínua. Uma variável aleatória que possui apenas partes discreta e contínua é conhecida como uma variável aleatória **mista**. Na prática, é pouco provável que surja uma variável aleatória singular. Portanto, quase todas as variáveis aleatórias são discretas, contínuas ou mistas.

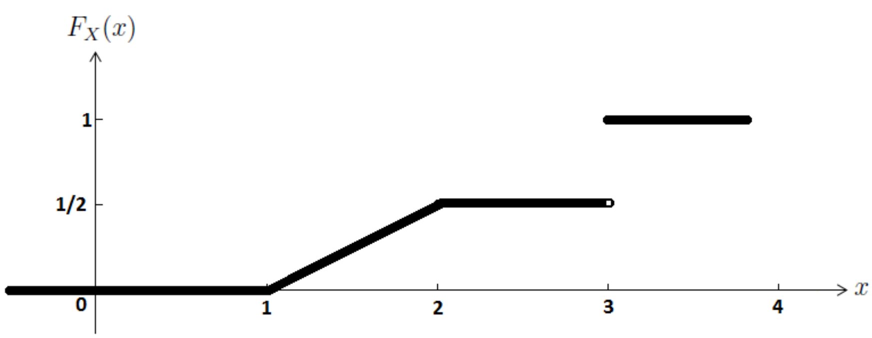

Figura 4.4: Gráfico de função de distribuição acumulada para uma variável aleatória mista

Exemplo 4.4.8: Exemplo de cálculo de probabilidades com uma variável aleatória mista. Seja

$$F_X(x) = \begin{cases} 0, & x < 1, \\ \frac{1}{4}(x-1), & 1 \leq x < 2, \\ \frac{1}{2}, & 2 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$$

(a) $P(X = 0.5) = F_X(0.5^+) - F_X(0.5^-) = 0 - 0 = 0.$

(b) $P(X = 1.5) = F_X(1.5^+) - F_X(1.5^-) = \frac{1}{4}(1.5-1) - \frac{1}{4}(1.5-1) = 0.$

(c) $P(X = 2) = F_X(2^+) - F_X(2^-) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$.

(d) $P(1 < X < 2) = F_X(2) - F_X(1) - P(X = 2) = \frac{2}{4} - 0 - \frac{1}{4} = \frac{1}{4}$.

(e) $P(1.5 \leq X \leq 2.5) = F_X(2.5) - F_X(1.5) + P(X = 1.5) = \frac{2}{4} - \frac{1}{4}(0.5) + 0 = 0.35$.

(f) $P(2 \leq X \leq 2.5) = F_X(2.5) - F_X(2) + P(X = 2) = \frac{2}{4} - \frac{2}{4} + (\frac{2}{4} - \frac{1}{4}) = \frac{1}{4}$.

(g) $P(2 < X \leq 3) = F_X(3) - F_X(2) = 1 - \frac{2}{4} = \frac{1}{2}$.

(h) $P(2 \leq X \leq 3) = F_X(3) - F_X(2) - (P(X = 3) - P(X = 2)) = 1 - \frac{2}{4} - ((1 - \frac{2}{4}) - (\frac{2}{4} - \frac{1}{4})) = \frac{1}{4}$.

(i) $P(X < 3.7) = F_X(3.7) - P(X = 3.7) = 1 - 0 = 1$.

(j) $P(X < 2) = F_X(2) - P(X = 2) = \frac{2}{4} - (\frac{2}{4} - \frac{1}{4}) = \frac{1}{4}$.

(k) $P(X \leq 1) = F_X(1) = \frac{1}{4}(1 - 1) = 0$.

(l) $P(X \leq 3) = F_X(3) = 1$. ■

4.5 Distribuições de caudas-pesadas

Definição 4.5.1: Uma variável aleatória X segue uma distribuição de caudas-pesadas se

$$\lim_{x \to \infty} e^{\lambda x} P(X > x) = \infty, \ \forall \lambda > 0.$$

Exemplo 4.5.2: Seja X tal que

$$f_X(x) = \alpha k^{\alpha} x^{-(\alpha+1)}, k \in \mathbb{R}^+, x \geq k, \alpha > 0.$$

Então X tem caudas-pesadas porque

$$F_X(x) = 1 - (\frac{k}{x})^{\alpha}, x \geq k \Rightarrow P(X > x) = (\frac{k}{x})^{\alpha}.$$

Esta é a distribuição de Pareto, que será vista no Capítulo 8. ■

A Tabela 4.2 mostra resultados para a distribuição de Pareto, quando $k = \alpha = 1$ e outra variável aleatória[2] de densidade $f_X(x) = e^{-x}, x > 0$, a qual não tem caudas-pesadas.

Observa-se que as probabilidades da variável que não tem caudas-pesadas rapidamente tendem a zero. Não são apenas variáveis aleatórias contínuas que têm caudas-pesadas. Segundo Fine (2006), a Zeta que é discreta também tem caudas-pesadas. Variáveis de caudas-pesadas têm sido usadas na modelagem de tráfego de redes de computadores porque, como o tráfego nas redes é variado (som, imagem, pacotes, dados, etc) qualquer que seja a variável observada, esta apresentará grande variabilidade o que implica no aumento da variância e consequentemente probabilidades altas nas caudas da distribuição (Adler, Feldman, Taqqu, (1998)).

[2]a Exponencial que também será vista no Capítulo 8.

Tabela 4.2: Resultados: Pareto *versus* Exponencial.

x	$P(X > x) = e^{-x}$	$P(X > x) = x^{-1}$
2	0.1353352...	0.5
5	0.0067397...	0.2
10	0.0000453...	0.1

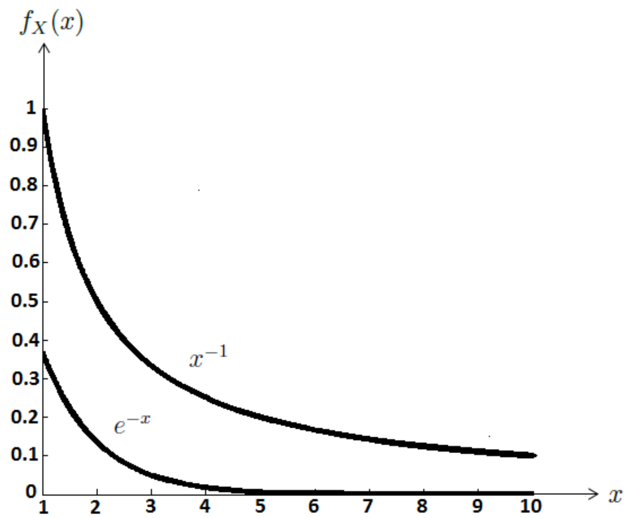

Figura 4.5: Comparação entre distribuição Exponencial e distribuição de Pareto

4.6 Funções de Variáveis Aleatórias

Muitas vezes é dada a distribuição de probabilidade que descreve o comportamento de uma variável aleatória X definida no espaço mensurável $(\Omega, \mathcal{A})$, mas o interesse é na descrição de uma função $Y = H(X)$. Por exemplo, X pode ser uma mensagem enviada em um canal de telecomunicações e Y ser a mensagem recebida.

Uma pergunta inicial é: se X é uma variável aleatória $\sqrt{X}$, $\log X$, X^2, $2X-3$ são variáveis aleatórias? Se sim, (o que é verdade), sendo conhecida a distribuição de probabilidade de X, como esse fato pode ser usado para encontrar a lei de probabilidade de $\sqrt{X}$, $\log X$, X^2 ou $2X-3$?

O problema é determinar $P(Y \in C)$, onde C é um evento Boreliano. Para determinar essa probabilidade, a imagem inversa da função H é fundamental, ou seja, a probabilidade do evento $\{Y \in C\}$ será por definição igual a probabilidade do evento $\{X \in H^{-1}(C)\}$, onde $H^{-1}(C) = \{x \in \mathbb{R} : H(x) \in C\}$. Para que esta probabilidade esteja bem definida, é preciso restringir H tal que $H^{-1}(C)$ seja um evento Boreliano para todo C Boreliano, caso contrário não é possível determinar $P(\{X \in H^{-1}(C)\})$; uma função que satisfaz esta condição é conhecida como *mensurável com respeito a $\mathcal{B}$*. Note que Y também pode ser vista como uma função do espaço amostral Ω, $Y(\omega) = H(X(\omega))$ para todo $\omega \in \Omega$. Vista dessa maneira Y é uma variável aleatória definida em $(\Omega, \mathcal{A})$, pois para todo Boreliano

C, $Y^{-1}(C) = X^{-1}(H^{-1}(C))$ e como por suposição $H^{-1}(C)$ é Boreliano porque X é uma variável aleatória, tem-se que $X^{-1}(H^{-1}(C)) \in \mathcal{A}$ e portanto satisfaz a definição de uma variável aleatória. A figura abaixo exibe os espaços mensuráveis e as transformações entre eles.

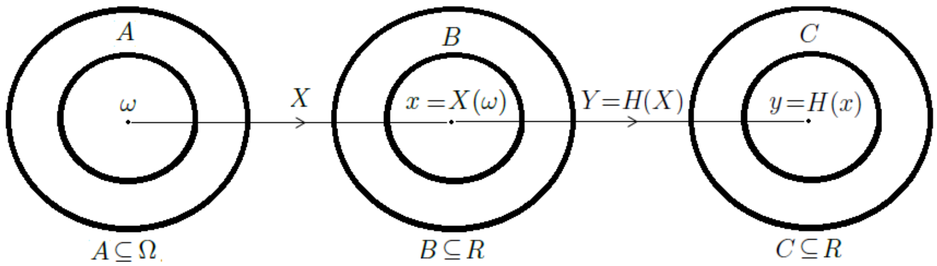

Figura 4.6: Mapeamento realizado por funções de variáveis aleatórias

Seja $A = \{\omega \in \Omega : X(\omega) \in B\}$. Portanto, como já mencionado anteriormente, a probabilidade induzida pela variável aleatória X é tal que

$$P_X(B) = P(X^{-1}(B)) = P(A).$$

De forma similar, sendo

$$B = Y^{-1}(C) = \{x \in \mathbb{R} : H(x) \in C\}$$

então,

$$P_Y(C) = P_{H(X)}(C) = P_X(\{x \in \mathbb{R} : H(x) \in C\}) = P(\{\omega \in \Omega : H(X(\omega)) \in C\}),$$

e assim,

$$P_Y(C) = P_X(Y^{-1}(C)).$$

Logo,

$$P_Y(C) = P_X(B) = P(A).$$

A função H da variável aleatória X define uma variável aleatória no espaço de probabilidade $(\mathbb{R}, \mathcal{B}, P_Y)$, onde a medida de probabilidade P_Y é induzida pela variável aleatória $Y = H(X)$. P_Y está bem definida pois

$$Y^{-1}(C) = B \in \mathcal{B},$$

o que mostra que a imagem inversa do conjunto mensurável C é o conjunto mensurável B. Adicionalmente, P_Y satisfaz os axiomas K2, K3, e K5′ porque:

(K2)

$$P_Y(C) = P_X(Y^{-1}(C)) = P_X(B) = P(X^{-1}(B)) = P(A) \geq 0.$$

(K3)

$$P_Y(\mathbb{R}) = P_X(Y^{-1}(\mathbb{R})) = P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1.$$

(K5′) Sejam $C_1, C_2, \ldots,$ Borelianos tais que $C_i \cap C_j = \emptyset$, para todo $i \neq j$ e $Y^{-1}(C_n) = B_n$. Então,

$$\begin{aligned} P_Y(\cup_n C_n) &= P_X(Y^{-1}(\cup_n C_n)) \\ &= P_X(\cup_n B_n) \\ &= \sum_n P_X(B_n) \\ &= \sum_n P_X(Y^{-1}(C_n) \\ &= \sum_n P_Y(C_n). \end{aligned}$$

Vale ressaltar que se X for uma variável aleatória discreta, $Y = H(X)$ também o será. Os exemplos a seguir ilustram como calcular a distribuição de probabilidade de uma função de variável aleatória. Ressalta-se a importância fundamental da função de distribuição acumulada, F, e de gráficos para visualizar as regiões C e B.

Exemplo 4.6.1: X, discreta; $Y = H(X)$, discreta. Admita-se que X tenha os valores possíveis $1, 2, 3, \ldots$ e que $P(X = n) = (1/2)^n$. Seja $Y = 1$ se X for par e $Y = -1$ se X for ímpar. Determinar a distribuição de Y.
Solução: Então,

$$P(Y = 1) = \sum_{n=1}^{\infty} (1/2)^{2n} = \sum_{n=1}^{\infty} (1/4)^n = \frac{1/4}{1 - 1/4} = 1/3.$$

Consequentemente,

$$P(Y = -1) = 1 - P(Y = 1) = 2/3.$$

■

De modo geral, suponha que X assume os valores $x_1, x_2, \ldots$ e que H uma função real tal que $Y = H(X)$ assume os valores $y_1, y_2, \ldots$. Agrupando os valores que X assume de acordo os valores de suas imagens quando se aplica a função H, ou seja, denotando por $x_{i1}, x_{i2}, x_{i3}, \ldots$ os valores de X tal que $H(x_{ij}) = y_i$ para todo j, tem-se que

$$P(Y = y_i) = P(X \in \{x_{i1}, x_{i2}, x_{i3}, \ldots\}) = \sum_{j=1}^{\infty} P(X = x_{ij}) = \sum_{j=1}^{\infty} p_X(x_{ij}),$$

ou seja, para calcular a probabilidade do evento $\{Y = y_i\}$, acha-se o evento equivalente em termos de X, isto é, todos os valores x_{ij} de X tal que $H(x_{ij}) = y_i$ e somam-se as probabilidades de X assumir cada um desses valores. ■

Exemplo 4.6.2: X**, discreta;** $Y = H(X)$**, discreta.** Seja X como no exemplo anterior e $Y = H(X) = X^2$. Determinar a lei de Y.
Solução: O contradomínio da variável Y, R_Y, e as respectivas probabilidades são:

$$R_Y = \{1, 4, \ldots, n^2, \ldots\},$$

$$P(Y = 1) = P(X = 1) = \frac{1}{2},$$

$$P(Y = 4) = P(X = 2) = \frac{1}{4},$$

$$\ldots$$

$$P(Y = n^2) = P(X = n) = \frac{1}{2^n}.$$

$$\ldots$$

■

Exemplo 4.6.3: X**, contínua;** $Y = H(X)$**, discreta.** Seja $f_X(x) = 2x$, $0 < x < 1$ e $Y = H(X)$ definida por $Y = 0$ se $X < \frac{1}{3}$, $Y = 1$, se $\frac{1}{3} \leq X < \frac{2}{3}$ e $Y = 2$, se $X \geq \frac{2}{3}$. Determinar a distribuição de Y.
Solução: Em termos de eventos equivalentes tem-se que:

$$C_1 = \{Y = 0\} \equiv B_1 = \{X < \frac{1}{3}\},$$

$$C_2 = \{Y = 1\} \equiv B_2 = \{\frac{1}{3} \leq X < \frac{2}{3}\},$$

$$C_3 = \{Y = 2\} \equiv B_3 = \{X \geq \frac{2}{3}\}.$$

Logo,

$$P(Y = 0) = P(X < \frac{1}{3}) = \int_0^{\frac{1}{3}} 2x dx = \frac{1}{9},$$

$$P(Y = 1) = P(\frac{1}{3} < X \leq \frac{2}{3}) = \int_{\frac{1}{3}}^{\frac{2}{3}} 2x dx = \frac{3}{9},$$

$$P(Y = 2) = P(X \geq \frac{2}{3}) = \int_{\frac{2}{3}}^{1} 2x dx = \frac{5}{9}.$$

■

Exemplo 4.6.4: X**, contínua;** $Y = H(X)$**, contínua.** Seja a densidade de X como no exemplo anterior e $Y = H(X) = e^{-X}$. Determinar a distribuição de Y.
Solução: O evento onde a densidade de X é não nula é $B = \{0 < X < 1\}$.

Portanto, a densidade de Y está concentrada em $C = \{y = H(x) : x \in B\} = \{e^{-1} < y < 1\}$ e

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(e^{-X} \leq y) \\ &= P(-X \leq \ln y) \\ &= P(X \geq -\ln y) \\ &= \int_{-\ln y}^{1} 2x dx \\ &= 1 - (-\ln y)^2 \Rightarrow f_Y(y) = \frac{-2\ln y}{y}. \end{aligned}$$

Logo,

$$f_Y(y) = \begin{cases} \frac{-2\ln y}{y}, & y \in (e^{-1}, 1), \\ 0, & y \notin (e^{-1}, 1). \end{cases}$$

■

Exemplo 4.6.5: Se $f_X(x) = 1$, $0 < x < 1$, e zero para quaisquer outros valores, qual a distribuição de $Y = -\log(X)$?
Solução: Como

$$0 < Y < \infty \Leftrightarrow 0 < X < 1$$

e $P(0 < X < 1) = 1$, tem-se $F_Y(y) = 0, y \leq 0$. Se $y > 0$, então

$$P(Y \leq y) = P(-\log(X) \leq y) = P(X \geq e^{-y}) = 1 - e^{-y},$$

ou seja, $Y \sim Exp(1)$, isto é, uma Exponencial (que será vista depois) de parâmetro 1. ■

No exemplo a seguir X é contínua e $H(X)$ é contínua. A ênfase deste exemplo é mostrar o cuidado na busca dos eventos equivalentes.

Exemplo 4.6.6: Seja $f_X(x) = \frac{1}{3}x^2$, $-1 < x < 2$ e zero para quaisquer outros valores de x. Encontrar a função densidade da variável aleatória $Y = X^2$.
Solução: Portanto, como pode ser visto na figura abaixo,

$$-1 < x < 1 \Rightarrow 0 < y < 1$$

e

$$1 \leq x < 2 \Rightarrow 1 \leq y < 4.$$

Então,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) + P(X = -\sqrt{y}) \\ &= \begin{cases} F_X(\sqrt{y}) - F_X(-\sqrt{y}), & 0 < y < 1, \\ F_X(\sqrt{y}), & 1 \leq y < 4. \end{cases} \end{aligned}$$

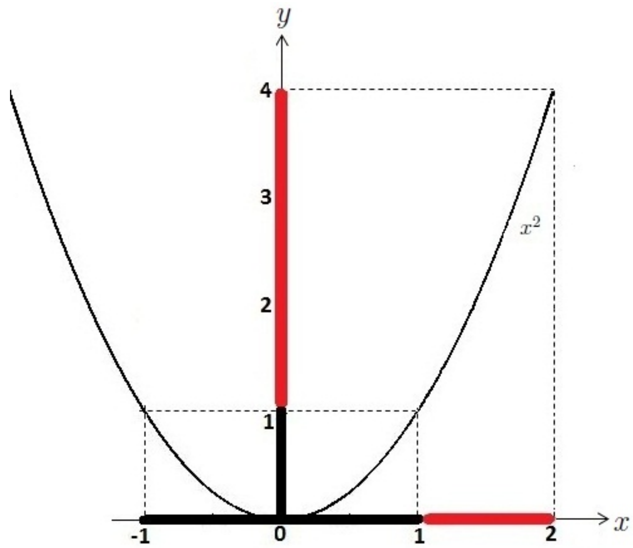

Figura 4.7: Mapeamento realizado por funções de variáveis aleatórias

Portanto,

$$f_Y(y) = \begin{cases} \frac{\sqrt{y}}{3}, & y \in (0,1), \\ \frac{\sqrt{y}}{6}, & y \in [1,4). \\ 0, & y \notin (0,4). \end{cases}$$

■

No caso de X e Y serem contínuas, tem-se o teorema seguinte.

Teorema 4.6.7: *Seja H uma função diferenciável, crescente ou decrescente em um dado intervalo I, $H(I)$ o contradomínio de H, H^{-1} a função inversa de H e X uma variável aleatória contínua com função densidade $f_X(x) > 0$, se $x \in I$ e $f_X(x) = 0$, se $x \notin I$. Então, $Y = H(X)$ tem função densidade de probabilidade dada por:*

$$f_Y(y) = \begin{cases} 0, & y \notin H(I), \\ f_X(H^{-1}(y))|\frac{dH^{-1}(y)}{dy}|, & y \in H(I). \end{cases}$$

Prova:

(a) H é crescente. Logo, H^{-1} também é crescente em I. Portanto,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(H(X) \leq y) \\ &= P(X \leq H^{-1}(y)) \\ &= F_X(H^{-1}(y)). \end{aligned}$$

Logo,

$$\frac{d}{dy}F_Y(y) = \frac{d}{dy}F_X(H^{-1}(y)) = \frac{dF_X(H^{-1}(y))}{dx}\frac{dx}{dy},$$

onde $x = H^{-1}(y)$.

Mas,

$$\frac{d}{dy}F_Y(y) = F'_Y(y) = f_Y(y).$$

Portanto,

$$\frac{dF_X(H^{-1}(y))}{dx}\frac{dx}{dy} = F'_X(H^{-1}(y))\frac{dH^{-1}(y)}{dy}.$$

Logo,

$$f_Y(y) = f_X(H^{-1}(y))\frac{dH^{-1}(y)}{dy}, y \in H(I).$$

(b) H é decrescente em I. Então H^{-1} também é decrescente em I. Logo,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(H(X) \leq y) \\ &= P(X \geq H^{-1}(y)) \\ &= 1 - F_X(H^{-1}(y)) + P(X = H^{-1}(y)) \\ &= 1 - F_X(H^{-1}(y)). \end{aligned}$$

Porque $P(X = H^{-1}(y)) = 0$ e seguindo o procedimento visto em (a),

$$F'_Y(y) = -F'_X(H^{-1}(y))\frac{dH^{-1}(y)}{dy}$$

e assim

$$f_Y(y) = -f_X(H^{-1}(y))\frac{dH^{-1}(y)}{dy}, y \in H(I).$$

∎

Também pode-se utilizar o método acima em outros casos em que a função H não seja nem crescente nem decrescente em I. Para tanto suponha que I possa ser dividido em uma quantidade enumerável $I_1, I_2, I_3, \ldots$ de subintervalos tal que H seja crescente ou decrescente em cada um deles, $P_X(I_j \cap I_k) = 0$ e $H(I_j) = H(I_k)$ para todo $j \neq k$. Neste caso, seja H_j^{-1} a função inversa de H restrita ao subintervalo I_j. Portanto,

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(H(X) \leq y) \\
&= \sum_{j:H_j^{-1}\text{é crescente}} P(X \leq H_j^{-1}(y)) + \sum_{j:H_j^{-1}\text{é decrescente}} P(X \geq H_j^{-1}(y)).
\end{aligned}
$$

Logo, pelos resultados anteriores,

$$f_Y(y) = \sum_j f_X(H_j^{-1}(y))|\frac{d}{dy}H_j^{-1}(y)|, y \in H(I).$$

Exemplo 4.6.8: Seja X com densidade $f_X(x)$ e $Y = X^2$. Então

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(X^2 \leq y) \\
&= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y}) + P(X = -\sqrt{y}) \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y}),
\end{aligned}
$$

porque $P(X = -\sqrt{y}) = 0$. Logo,

$$\frac{d}{dy}F_Y(y) = \frac{d}{dy}(F_X(\sqrt{y}) - F_X(-\sqrt{y})) = \frac{d}{dy}F_X(\sqrt{y}) - \frac{d}{dy}F_X(-\sqrt{y}).$$

Mas,

$$\frac{d}{dy}F_Y(y) = f_Y(y),$$

$$\frac{d}{dy}F_X(\sqrt{y}) = \frac{dF_X(\sqrt{y})}{dx_1}\frac{dx_1}{dy},\ x_1 = \sqrt{y},$$

$$\frac{dF_X(\sqrt{y})}{dx_1} = f_X(\sqrt{y}),$$

$$\frac{dx_1}{dy} = \frac{1}{2\sqrt{y}},$$

$$\frac{d}{dy}F_X(-\sqrt{y}) = \frac{dF_X(-\sqrt{y})}{dx_2}\frac{dx_2}{dy},\ x_2 = -\sqrt{y},$$

$$\frac{dF_X(-\sqrt{y})}{dx_2} = f_X(-\sqrt{y}),$$

$$\frac{dx_2}{dy} = -\frac{1}{2\sqrt{y}}.$$

Logo,

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y})), & y \geq 0, \\ 0, & y < 0, \end{cases}$$

Alternativamente, poderia ter sido usado o procedimento descrito anteriormente e particionar $\mathbb{R}$ nos subintervalos $I_1 = (-\infty, 0]$ e $I_2 = [0, +\infty)$. Note que $P_X(I_1 \cap I_2) = 0$, $H(I_1) = H(I_2) = [0, +\infty)$, $H_1^{-1}(y) = -\sqrt{y}$ e $H_2^{-1}(y) = \sqrt{y}$. Portanto,

$$f_Y(y) = f_X(-\sqrt{y})\frac{1}{2\sqrt{y}} + f_X(\sqrt{y})\frac{1}{2\sqrt{y}}, y \geq 0.$$

■

Exemplo 4.6.9: Este exemplo mostra diversos métodos de resolver um problema sobre como calcular a densidade, ou distribuição acumulada, para uma dada variável aleatória. Seja

$$f_X(x) = \begin{cases} \frac{1}{2}, & -1 < x < 1, \\ 0, & x \notin (-1, 1), \end{cases}$$

e $Y = H(X) = X^2$. O objetivo é calcular $f_Y(y)$.

Como $-1 < x < 1$ e $y = x^2$, então $0 < y < 1$.

(a)

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) + P(X = -\sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

Mas,

$$\frac{dF_Y(y)}{dy} = f_Y(y)$$

e

$$\frac{dF_Y(y)}{dy} = \frac{dF_X(\sqrt{y})}{d(\sqrt{y})} \times \frac{d(\sqrt{y})}{dy} - \frac{dF_X(-\sqrt{y})}{d(-\sqrt{y})} \times \frac{d(-\sqrt{y})}{dy}.$$

Logo,

$$\begin{aligned} f_Y(y) &= \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) \\ &= \frac{1}{2\sqrt{y}}\frac{1}{2} + \frac{1}{2\sqrt{y}}\frac{1}{2} \\ &= \frac{1}{2\sqrt{y}}, 0 < y < 1. \end{aligned}$$

(b)

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(X^2 \leq y) \\
&= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y}) + P(X = -\sqrt{y}) \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\
&= \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx \\
&= \frac{1}{2}(\sqrt{y} + \sqrt{y}) \\
&= \sqrt{y}.
\end{aligned}$$

Portanto, como $f_Y(y) = F'_Y(y)$, então,

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

(c) Usando o Teorema 4.6.7.

$$f_Y(y) = \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y})) = \frac{1}{2\sqrt{y}}(\frac{1}{2} + \frac{1}{2}) = \frac{1}{2\sqrt{y}}, 0 < y < 1.$$

(d) Calculando inicialmente $F_X(x)$.

$$F_X(x) = \begin{cases} 0, & x < -1, \\ \frac{x+1}{2}, & -1 \leq x < 1, \\ 1, & x \geq 1. \end{cases}$$

Mas,

$$F_X(\sqrt{y}) = \frac{\sqrt{y}+1}{2},$$

$$F_X(-\sqrt{y}) = \frac{-\sqrt{y}+1}{2}.$$

Logo,

$$F_Y(y) = \begin{cases} 0, & y < 0, \\ \frac{\sqrt{y}+1}{2} - \frac{-\sqrt{y}+1}{2} = \sqrt{y}, & 0 \leq y < 1, \\ 1, & y \geq 1. \end{cases}$$

Portanto, derivando,

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

■

4.7 Aprendendo um pouco mais

O exemplo de uma variável aleatória singular é a função de Cantor, cuja construção segue-se.

Exemplo 4.7.1: Seja

$$F_0(x) = \begin{cases} 0, & x < 0, \\ 1, & x > 1. \end{cases}$$

Dividindo-se o intervalo $(0,1)$ nos três subintervalos $(0, \frac{1}{3})$, $(\frac{1}{3}, \frac{2}{3})$ e $(\frac{2}{3}, 1)$ e considerando-se como valor de F em $(\frac{1}{3}, \frac{2}{3})$ a média dos valores de F_0 fora de $(0,1)$, isto é, $\frac{0+1}{2} = \frac{1}{2}$, obtém-se $F_1(x)$:

$$F_1(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{2}, & \frac{1}{3} < x < \frac{2}{3}, \\ 1, & x > 1. \end{cases}$$

Cada terço do intervalo $(0,1)$ sendo dividido em três partes equivale a dividir $(0,1)$ em nove partes. Para o intervalo $(\frac{1}{9}, \frac{2}{9})$, o valor da F é $\frac{0+\frac{1}{2}}{2} = \frac{1}{4}$; para o intervalo $(\frac{7}{9}, \frac{8}{9})$, o valor da F é $\frac{\frac{1}{2}+1}{2} = \frac{3}{4}$.

Este processo constrói uma sequência de funções $F_n(x)$, $n = 1, 2, \cdots$, cuja função limite, $F(x)$, satisfaz às propriedades F1, F2, F3. Além disso, F é uma função contínua onde o conjunto de pontos onde a função cresce tem comprimento nulo. Portanto, F é uma função de distribuição, entretanto não é nem discreta, nem contínua, é uma variável aleatória singular.

■

Pode ser visto (James, 1981) que toda variável aleatória é uma combinação dos três tipos: discreta, contínua e singular; entretanto, as variáveis aleatória que são comuns no mundo real ou são discretas, ou contínuas, ou uma combinação entre esses dois tipos (mistas).

O exemplo a seguir mostra como decompor F em suas partes discreta, contínua e singular.

Exemplo 4.7.2: Suponha que X tenha densidade $f_X(x) = 1$ para $0 < x < 1$ e 0 para quaisquer outros valores. Seja $Y = min\{X, 1/2\}$. Note que

$$F_Y(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \leq x < 1/2, \\ 1, & x \geq 1/2. \end{cases}$$

F_Y tem apenas um salto em $x = 1/2$ e $p_1 = 1/2$. Logo, $F_d(x) = 0$ se $x < 1/2$ e $F_d(x) = 1/2$ se $x \geq 1/2$. Diferenciando F_Y, tem-se

$$F'_Y(x) = \begin{cases} 0, & x < 0 \vee x > 1/2, \\ 1, & 0 < x < 1/2. \end{cases}$$

Portanto,

$$F_{ac}(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} 0, & x < 0, \\ x, & 0 \leq x \leq 1/2, \\ 1/2, & x > 1/2. \end{cases}$$

Como $F_d + F_{ac} = F_Y$, tem-se que $F_s(x) = 0$, $\forall x \in \mathbb{R}$ e não há parte singular. ■

4.8 Exercícios

4.1 Seja a variável aleatória X tal que

$$f_X(x) = \begin{cases} c(1-x^2), & -1 < x < 1, \\ 0, & x \notin (-1,1). \end{cases}$$

(a) Qual é o valor de c?

(b) Qual é a função de distribuição acumulada de X?

4.2 Seja X tal que

$$f_X(x) = \begin{cases} c(4x-2x^2), & 0 < x < 2, \\ 0, & x \notin (0,2). \end{cases}$$

(a) Qual é o valor de c?

(b) Determine $P(\frac{1}{2} < X < \frac{3}{2})$ usando $f_X(x)$ e $F_X(x)$.

4.3 Resolva este exercício usando um *software* adequado. Seja a função de distribuição acumulada da variável aleatória X,

$$F_X(x) = \begin{cases} 0, & x < 0, \\ (2/\pi)\sin^{-1}(\sqrt{x}), & 0 \le x < 1, \\ 1, & x \ge 1. \end{cases}$$

Faça o gráfico de $F_X(\cdot)$. Determine a função densidade de probabilidade e faça seu gráfico.

4.4 Para cada uma das funções abaixo, faça seu gráfico; verifique se é uma função densidade de probabilidade para uma dada variável aleatória X. Se for, encontre a função de distribuição acumulada e faça seu gráfico.

(a) $f_X(x) = 6x(1-x)$, $0 \le x \le 1$.

(b)

$$f_X(x) = \begin{cases} 1+x, & -1 \le x \le 0, \\ 1-x, & 0 \le x \le 1, \\ 0, & \text{quaisquer outros valores.} \end{cases}$$

(c)

$$f_X(x) = \begin{cases} x/2, & 0 \le x \le 1, \\ 1/2, & 1 \le x \le 2, \\ -x/2+3/2, & 2 \le x \le 3, \\ 0, & \text{quaisquer outros valores.} \end{cases}$$

4.5 Uma variável aleatória contínua X tem função densidade $f_X(x) = \alpha e^{-\alpha x}$, $x > 0$ e $\alpha > 0$.

(a) Determine a função de distribuição acumulada de X.

(b) Calcule as seguintes probabilidades usando a função encontrada no item anterior:

(b1) $P(X \leq 3)$.

(b2) $P(X > 2)$.

(b3) $P(X < -1)$.

(b4) $P(X > -1)$.

4.6 Um ponto é escolhido ao acaso sobre uma reta de comprimento L. Qual é a probabilidade de que a razão do segmento mais curto para o mais longo seja menor que $1/2$?

4.7 Uma variável aleatória X tem densidade $f_X(\cdot)$ dada por

$$f_X(x) = \begin{cases} \alpha x, & 0 \leq x < 0.5, \\ \alpha(1-x), & 0.5 \leq x < 1, \\ 0, & \text{quaisquer outros valores.} \end{cases}$$

(a) Determine o valor da constante α.

(b) Sejam os eventos $A = \{X < 0.5\}$, $B = \{X > 0.5\}$ e $C = \{0.25 < X < 0.75\}$.

(b1) Calcule $P(A \mid B)$.

(b2) Verifique se A, B e C são mutuamente independentes.

4.8 Um motorista tem que, obrigatoriamente, passar em 4 (e somente 4) semafóros para alcançar seu destino. Em cada um deles, independentemente, a probabilidade do carro parar é p. Seja uma variável aleatória X, definida como sendo o número de semáforos que o carro passa antes de parar pela primeira vez. Estabeleça a distribuição de probabilidade de X. Prove que a expressão encontrada é realmente uma distribuição de probabilidade.

4.9 Em um jogo de dados, A paga R\$20,00 a B e lança três dados honestos. Se sair a face 1 em no máximo um dos dados, A ganha R\$20,00 de B; se sair face 1 em dois dados apenas, A ganha R\$50,00; se sair face 1 nos três dados, A ganha R\$80,00. Determine a distribuição de probabilidade do lucro líquido por jogada.

4.10 Seja uma variável aleatória contínua X, com função de densidade $f_X(x) = \alpha e(-\lambda \mid x \mid)$, com $x \in \mathbb{R}$ e $\alpha > 0$.

(a) Determine a constante α.

(b) Esboçe o gráfico de $f_X(x)$.

(c) Determine $F_X(x)$.

(d) Determine m tal que $P(X \leq m) = P(X > m)$.

4.11 Suponha que a função de distribuição acumulada para uma variável aleatória X, $F_X(\cdot)$, fosse definida por $F_X(x) = P(X < x)$. Usando esta definição determine as seguintes probabilidades:

(a) $P(X \leq x)$.

(b) $P(a \leq X \leq b)$.

(c) $P(a \leq X < b)$.

(d) $P(a < X < b)$.

Sugestões: $(-\infty, a] = (-\infty, a) \cup \{a\}$, $(-\infty, a] \cup (a, b) = (-\infty, b)$.

4.12 Seja $f_U(u) = e^{-u}, u \geq 0$. Mostre que f é uma função densidade. Encontre $\int_0^\infty u f_U(u) du$.

4.13 Suponhamos que dez cartas estejam numeradas de 1 até 10. Das dez cartas, retira-se uma de cada vez, ao acaso e sem reposição, até retirar-se o primeiro número par. Conta-se o número de retiradas necessárias. Exiba um bom modelo probabilístico para este experimento.

4.14 Seja X uma variável aleatória com densidade

$$f_X(x) = \begin{cases} cx^2, & -1 \leq x \leq 1, \\ 0, & \text{caso contrário.} \end{cases}$$

(a) Determine o valor da constante c.

(b) Determine a função de distribuição acumulada e esboçe seu gráfico.

(c) Ache o valor α tal que $F_X(\alpha) = 1/4$. (α é o primeiro *quartil* da distribuição de X.)

(d) Ache o valor β tal que $F_X(\beta) = 1/2$. (β é a *mediana* da distribuição de X.)

4.15 Uma variável aleatória X tem função distribuição

$$F_X(x) = \begin{cases} 1, & x > 1, \\ x^3, & 0 \leq x \leq 1, \\ 0, & x < 0. \end{cases}$$

Qual é a densidade de X?

4.16 Uma variável X tem função de distribuição

$$F_X(x) = \begin{cases} 0, & x < 0, \\ x^2/2, & 0 \leq x < 1, \\ 3/4, & 1 \leq x < 2, \\ (1/4)(x+1), & 2 \leq x < 3, \\ 1, & x \geq 3. \end{cases}$$

Determine o seguinte:

(a) $P(X = 1/2)$;

(b) $P(X = 1)$;

(c) $P(X < 1)$;

(d) $P(X \leq 1)$;

(e) $P(X > 2)$;

(f) $P(1/2 < X < 5/2)$.

4.17 Calcule

(a) $P(X > 2)$;

(b) $P(X \leq 0)$;

(c) $P(X = 0)$;

(d) $P(X < 0)$;

(e) $P(X \geq 0.5)$.

para uma variável X que tem função de distribuição

$$F_X(x) = \begin{cases} 1 - 0.75e^{-x}, & \text{se } x \geq 0, \\ 0, & \text{se } x < 0. \end{cases}$$

4.18 Seja a probabilidade da variável aleatória X definida por $P(A) = \int_A f_X(x)dx$, onde $f_X(x) = cx/9$, para $0 < x < 3$. Sejam $A_1 = \{x \mid 0 < x < 1\}$ e $A_2 = \{x \mid 2 < x < 3\}$. Calcule

(a) o valor da constante c,

(b) $P(A_1)$,

(c) $P(A_2)$,

(d) $P(A_1 \cup A_2)$,

(e) $P(A_1 \mid A_2)$.

4.19 Coloque V ou F nas sentenças abaixo:

(a) Uma variável aleatória X só assume valores no intervalo $[0, 1]$. ( )

(b) Se X é uma variável aleatória contínua, então X também é uma variável aleatória discreta. ( )

(c) Se X é uma variável aleatória discreta então X não pode ser contínua. A recíproca é que é verdadeira. ( )

(d) Se X é uma variável aleatória contínua, $F_X(x) = \int_{-\infty}^{x} f_X(s)ds$. ( )

(e) Se X é uma variável aleatória contínua, $f_X(x) = \frac{d}{dx}F_X(x)$. ( )

(g) $\lim_{x\to+\infty} F_X(x) = 0$. ( )

(h) $P(X \in A) = \int_A F_X(x)dx$. ( )

(i) $P(X \in A) = \int_A f_X(x)dx$. ( )

4.20 Foguetes são lançados até que o primeiro lançamento bem sucedido tenha ocorrido. Se isso não ocorrer até 5 tentativas, o experimento é suspenso e o equipamento inspecionado. Admita que exista uma probabilidade constante de 0.8 de haver um lançamento bem sucedido e que os sucessivos lançamentos sejam independentes. Suponha que o custo do primeiro lançamento seja k dólares, enquanto os lançamentos subsequentes custam $k/3$ dólares. Sempre que ocorre um lançamento bem sucedido, uma certa quantidade de informação é obtida, a qual pode ser expressa como um ganho financeiro de c dólares. Seja T o custo líquido desse experimento. Estabeleça a distribuição de probabilidade de T.

4.21 Determine a densidade de $Y = (b-a)X + a$, onde $f_X(x) = 1$, se $0 < x < 1$ e zero para quaisquer outros valores.

4.22 Se X tem densidade $f_X(x) = e^{-|x|}/2,\ -\infty < x < +\infty$, qual é a distribuição de $Y =| X |$?

4.23 Uma variável aleatória X tem uma densidade de probabilidade $f_X(x)$. Encontre a função densidade de probabilidade da variável aleatória $Y = aX + b$, onde a e b são constantes.

4.24 Uma variável aleatória X tem uma densidade de probabilidade $f_X(x)$. Qual a função densidade de probabilidade da variável aleatória $Y =| 1 - X |$?

4.25 Uma variável aleatória contínua X tem uma densidade de probabilidade $f_X(x)$. Considere a variável $Y = -X$. Encontre sua função densidade $f_Y(y)$.

4.26 Uma variável aleatória contínua X tem uma densidade de probabilidade $f_X(x)$. Encontre a função densidade $f_Y(y)$ do seu módulo, $Y =| X |$.

4.27 Uma variável aleatória X tem uma função distribuição $F_X(x)$, e uma variável aleatória Y relaciona-se com X por $Y = 2 - 3X$. Encontre a função distribuição $F_Y(y)$ da variável aleatória Y.

4.28 Dada uma variável aleatória contínua X com função densidade $f_X(x)$, encontre a distribuição da variável aleatória

$$Y = \text{sinal de } X = \begin{cases} +1, & X > 0, \\ 0, & X = 0, \\ -1, & X < 0. \end{cases}$$

4.29 Uma variável aleatória X tem uma densidade de probabilidade correspondente a reta que passa pelos pontos $(-1, 0)$ e $(1, 1)$, para $x \in (-1, 1)$, e zero fora. Uma variável aleatória Y é relacionada a X por $Y = 1 - X^2$. Encontre a função densidade $f_Y(y)$.

4.30 Uma variável aleatória X tem densidade $f_X(x) = 1$, no intervalo $(0, 1)$ zero fora. Uma variável aleatória Y tem um relacionamento funcional monotonicamente crescente com a variável X tal que $Y = \varphi(X)$. Encontre a função distribuição $F_Y(y)$ e a função densidade $f_Y(y)$.

4.31 Seja X uma variável aleatória tal que $P(| X - 1 |= 2) = 0$. Expresse $P(| X - 1 |\geq 2)$ em termos da função de distribuição F_X.

4.32 Seja $f_X(x) = \frac{1}{3}, \; -1 < x < 2$ e zero para quaisquer outros valores de X. Encontre a função densidade da variável aleatória $Y = X^2$.

4.33 Seja X tendo função probabilidade $f_X(x) = (\frac{1}{2})^x, \; x = 1, 2, \cdots$ e zero para quaisquer outros valores de X. Encontre a função probabilidade de $Y = X^3$.

4.34 Seja X tendo função probabilidade $f_X(x) = x^2/9, \; 0 < x < 3$, e zero para quaisquer outros valores. Encontre a função probabilidade de $Y = X^3$.

4.35 Seja X tendo função densidade $f_X(x) = 2xe^{-x^2}, \; 0 < x < \infty$, e zero para quaisquer outros valores. Determine a densidade de $Y = X^2$.

4.36 Seja X uma variável aleatória contínua com função densidade $f_X(x)$.

(a) Encontre a função densidade de $Y = X^2$.

(b) Se $f_X(x) = f_X(-x), \; \forall x$, simplifique a resposta encontrada em (a).

(c) Se $f_X(x) = 0$ quando $x \leq 0$, simplifique a resposta encontrada em (a).

4.37 Uma variável aleatória X tem função densidade probabilidade definida por:

$$f_X(x) = \begin{cases} c + x, & -1 \leq x < 0, \\ c - x, & 0 \leq x \leq 1, \\ 0, & \text{quaisquer outros casos.} \end{cases}$$

(a) Calcule o valor da constante c.

(b) Seja o evento $A = \{x \mid -0.5 \leq x \leq 0.5\}$. Compute $P(A)$.

(c) Encontre a função de distribuição acumulada de X, F_X, e usando a mesma calcule $P(X \leq 0.5)$.

(d) Suponha que uma variável Y assuma o valor 0 se X for negativa e 1, se X for positiva ou nula. Encontre a distribuição de probabilidade dessa variável.

Capítulo 5

Vetores Aleatórios e Funções

5.1 Introdução

Muitas vezes na vida real, o interesse é na descrição probabilística de mais de um característico numérico de um experimento aleatório. Por exemplo, na distribuição de alturas e pesos de indivíduos de uma certa classe. Para tanto é preciso estender a definição de variável aleatória para o caso multidimensional.

Definição 5.1.1: Seja $(\Omega, \mathcal{A}, P)$ um espaço de probabilidade. Uma função $\vec{X} : \Omega \to \mathbb{R}^n$ é chamada de um *vetor aleatório* se para todo evento B, Boreliano,[1] de $\mathbb{R}^n$, $\vec{X}^{-1}(B) \in \mathcal{A}$.

Dado um vetor aleatório $\vec{X}$, pode-se definir uma probabilidade induzida $P_{\vec{X}}$ no espaço mensurável $(\mathbb{R}^n, \mathcal{B}^n)$ da seguinte maneira: para todo $B \in \mathcal{B}^n$, define-se $P_{\vec{X}}(B) = P(\vec{X}^{-1}(B))$. Por definição de vetor aleatório, tem-se que $\vec{X}^{-1}(B) = A \in \mathcal{A}$, então $P_{\vec{X}}$ está bem definida.

5.2 Função de Distribuição Acumulada Conjunta

Para um vetor aleatório $\vec{X}$, uma maneira básica de descrever a probabilidade induzida $P_{\vec{X}}$ é utilizando sua *função de distribuição acumulada conjunta*.

Definição 5.2.1: A *função de distribuição acumulada conjunta* de um vetor aleatório $\vec{X}$, representada por $F_{\vec{X}}$ ou simplesmente por F, é definida por

$$F_{\vec{X}}(\vec{x}) = P(B_{\vec{x}}) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n), \forall \vec{x} \in \mathbb{R}^n.$$

A função de distribuição acumulada $F_{\vec{X}}$ satisfaz às seguintes propriedades:

[1] Um evento é Boreliano em $\mathbb{R}^n$ se pertence a menor σ-álgebra que contem todas regiões da seguinte forma: $B_{\vec{x}} = \{(X_1, X_2, \ldots, X_n) : X_i \leq x_i, 1 \leq i \leq n\}$.

(F1) Se $x_i \leq y_i$, $\forall i \leq n$, então $F_{\vec{X}}(\vec{x}) \leq F_{\vec{X}}(\vec{y})$.

$$x_i \leq y_i \, \forall i \leq n \Rightarrow B_{\vec{x}} \subseteq B_{\vec{y}} \Rightarrow P(B_{\vec{x}}) \leq P(B_{\vec{y}}) \Rightarrow F_{\vec{X}}(\vec{x}) \leq F_{\vec{X}}(\vec{y}).$$

(F2) $F(x_1, x_2, \ldots, x_n)$ é contínua a direita em cada uma das variáveis. Por exemplo, se $y_m \downarrow x_1$, então

$$F(y_m, x_2, \ldots, x_n) \downarrow F(x_1, x_2, \ldots, x_n), \text{ quando } m \to \infty.$$

(F3a) Se para algum $i \leq n$, $x_i \to -\infty$, então $B_{\vec{x}}$ decresce monotonicamente para o conjunto vazio $\emptyset$. Logo, pela continuidade monotï£¡nica de probabilidade,

$$\lim_{x_i \to -\infty} F_{\vec{X}}(\vec{x}) = 0.$$

(F3b) Se $x_i \to \infty$, então $B_{\vec{x}}$ cresce monotonicamente para o conjunto $\{X_1 \leq x_1, \ldots X_{i-1} \leq x_{i-1}, X_{i+1} \leq x_{i+1}, \ldots, X_n \leq x_n\}$, ou seja a restrição em X_i é removida. Então, pode-se escrever

$$\lim_{x_i \to \infty} F_{\vec{X}}(\vec{x}) = F_{X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n).$$

Portanto, a função de distribuição acumulada conjunta de $X_1, \ldots, X_{n-1}$ pode ser facilmente determinada da função de distribuição acumulada conjunta de $X_1, \ldots, X_n$ fazendo $x_n \to \infty$. Observe que *funções de distribuição acumuladas conjuntas de ordem maiores determinam as de ordem menores, mas o contrário não é verdadeiro*. Em particular,

$$\lim_{\vec{x} \to \infty} F_{\vec{X}}(\vec{x}) = 1.$$

A função de distribuição acumulada de X_i que se obtém a partir da função acumulada conjunta de $X_1, \ldots, X_n$ fazendo $x_j \to \infty$ para $j \neq i$ é denominada de *função de distribuição marginal* de X_i.

O próximo exemplo (James, 1996) mostra que para $n \geq 2$ as propriedades F1, F2, e F3 não são suficientes para que F seja uma função de distribuição.

Exemplo 5.2.2: Seja $F_0 : \mathbb{R}^2 \to \mathbb{R}$ uma função definida no plano tal que $F_0(x, y) = 1$ se $x \geq 0$, $y \geq 0$, e $x + y \geq 1$, e $F_0(x, y) = 0$, caso contrário. É claro que F1, F2, e F3 são satisfeitas, mas F_0 não é função de distribuição de nenhum vetor aleatório (X, Y), porque tem-se a seguinte contradição:

$$\begin{aligned} & 0 \leq P(0 < X \leq 1, 0 < Y \leq 1) \\ & = F_0(1,1) - F_0(1,0) - F_0(0,1) + F_0(0,0) = 1 - 1 - 1 + 0 = -1 \end{aligned}$$

O resultado acima vem de:

$$F_0(1, 1) = P(X \leq 1, Y \leq 1),$$

$$F_0(1,0) = P(X \leq 1, Y \leq 0),$$
$$F_0(0,1) = P(X \leq 0, Y \leq 1),$$
$$F_0(0,0) = P(X \leq 0, Y \leq 0),$$

Logo,

$$\begin{aligned} F_0(1,1) - F_0(1,0) &= P(X \leq 1, Y \leq 1) - P(X \leq 1, Y \leq 0) && (5.1) \\ &= P(\{X \leq 1, Y \leq 1\}) - P(\{X \leq 1, Y \leq 0\}) && (5.2) \\ &= P(\{X \leq 1, Y \leq 1\} - \{X \leq 1, Y \leq 0\}) && (5.3) \\ &= P(X \leq 1, 0 < Y \leq 1). && (5.4) \end{aligned}$$

A Fórmula (5.4) decorre de $P(B - A) = P(B) - P(A)$, quando $A \subseteq B$.

De forma similar

$$F_0(0,1) - F_0(0,0) = P(X \leq 0, Y \leq 1) - P(X \leq 0, Y \leq 0) = P(X \leq 0, 0 < Y \leq 1). \quad (5.5)$$

Por fim, de (5.4) e (5.5),

$$P(X \leq 1, 0 < Y \leq 1) - P(X \leq 0, 0 < Y \leq 1) = P(0 < X \leq 1, 0 < Y \leq 1).$$

■

5.2.1 Vetor Aleatório Discreto

$\vec{X}$ é um *vetor aleatório discreto* se existir um conjunto enumerável $\{\vec{x}_1, \vec{x}_2, \ldots, \}$ tal que $P(\vec{x} \in \{\vec{x}_1, \vec{x}_2, \ldots, \}) = 1$. Neste caso, define-se uma *função probabilidade de massa conjunta*, ou sua *distribuição de probabilidade conjunta* p,

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = p(x_1, x_2, \ldots, x_n) = p(\vec{x}_i).$$

Logo,

$$p(\vec{x}_i) \geq 0,$$
$$\sum_{i=1}^{\infty} p(\vec{x}_i) = 1.$$

5.2.2 Vetor Aleatório Contínuo

Seja $\vec{X} = (X_1, \ldots, X_n)$ um vetor aleatório e F sua função de distribuição acumulada conjunta. Se existe uma função $f(x_1, \ldots, x_n) \geq 0$ tal que

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(t_1, \ldots, t_n) dt_1 \ldots dt_n, \forall (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

então f é chamada de *densidade conjunta* das variáveis aleatórias $X_1, \ldots, X_n$, e neste caso, $\vec{X}$ é *contínuo*.

Similar ao caso unidimensional,

$$\frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1, \ldots, \partial x_n} = f(x_1, \ldots x_n).$$

5.3 Distribuições Marginais e Condicionais

Definição 5.3.1: A *função probabilidade de massa marginal* ou a *distribuição de probabilidade marginal* de X_i é

$$p_{X_i}(x_i) = \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_n} p(x_1, \ldots, x_n).$$

Definição 5.3.2: A *densidade marginal* de X_i é

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n.$$

A seguir será visto como calcular probabilidades condicionais envolvendo variáveis aleatórias.

Definição 5.3.3: Sejam X e Y variáveis aleatórias com distribuição de probabilidade conjunta $P(X = x_i, Y = y_j) = p(x_i, y_j)$, (i, j) pertencente ao contradomínio de (X,Y). Então, a *distribuição condicional* de X dada $Y = y_j$, $P(X = x \mid Y = y_j)$, é

$$P(X = x_i \mid Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p(x_i, y_j)}{p_Y(y_j)} = p_{X|Y}(x_i|y_j),\ p_Y(y_j) > 0. \tag{5.6}$$

O leitor pode fazer uma analogia com a definição de probabilidade condicional vista anteriormente. Facilmente observa-se que (5.6) é uma probabilidade:

(i) $P(X = x_i \mid Y = y_j) \geq 0$, porque é quociente de probabilidades.

(ii)

$$\begin{aligned} P(X \in \mathbb{R} \mid Y = y_j) &= \frac{P(X \in \mathbb{R}, Y = y_j)}{P(Y = y_j)} \\ &= \frac{P(Y = y_j)}{P(Y = y_j)} = 1. \end{aligned}$$

(iii)

$$\begin{aligned} P(\cup_{i=1}^{\infty} \{X = x_i\} \mid Y = y_j) &= \frac{P((\cup_{i=1}^{\infty} \{X = x_i\}) \cap \{Y = y_j\})}{P(Y = y_j)} \\ &= \frac{P(\cup_{i=1}^{\infty} (\{X = x_i\} \cap \{Y = y_j\}))}{P(Y = y_j)} \\ &= \frac{P(\cup_{i=1}^{\infty} \{X = x_i, Y = y_j\})}{P(Y = y_j)} \\ &= \frac{\sum_{i=1}^{\infty} P(X = x_i, Y = y_j)}{P(Y = y_j)} \\ &= \sum_{i=1}^{\infty} P(X = x_i \mid Y = y_j). \end{aligned}$$

Analogamente,

$$P(Y = y_j \mid X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)}.$$

Quando as variáveis aleatórias X e Y são contínuas, o fato de $P(Y = y) = 0$, $\forall y$ em (5.6) torna necessária a adição de um conceito novo na definição das probabilidades condicionais.

Para resolver este caso, será utilizado um argumento de limites. Suponha que o objetivo seja definir $P(X \leq x|Y = y)$, onde Y é uma variável contínua. Por exemplo, X poderia ser alturas de indivíduos e Y seus respectivos pesos; então $\{Y = y\}$ significa que o peso está fixo e $P(X \leq x|Y = y)$ implica em mensurar todas as alturas menores ou iguais a x para o peso fixo em y. Deste modo, suponha que exista um intervalo I de comprimento δ contendo y em seu interior. $P(X \leq x|Y = y)$ pode ser aproximada por

$$P(X \leq x|Y \in I) = \frac{P(X \leq x, Y \in I)}{P(Y \in I)},$$

esta probabilidade está bem definida desde que $P(Y \in I) > 0$. Caso $P(Y \in I) = 0$, para algum intervalo contendo y, a definição da probabilidade $P(X \leq x|Y = y)$ pode ser arbitrária, pois tal valor y nunca ocorrerá. Esta aproximação será tão melhor quanto menor for δ. Desta forma, pode-se definir $P(X \leq x|Y = y)$ como sendo o limite $P(X \leq x|Y \in I)$ quando δ tende a zero. Assumindo que (X, Y) possui densidade conjunta $f(x, y)$, tem-se:

$$P(X \leq x|Y = y) = \lim_{\delta \to 0} \frac{P(X \leq x, Y \in I)}{P(Y \in I)} = \lim_{\delta \to 0} \frac{\int_{-\infty}^{x} \int_{y \in I} f(x, y)dydx}{\int_{y \in I} f(y)dy}.$$

Supondo $f(x, y)$ contínua na região em que $y \in I$,

$$P(X \leq x|Y = y) = \lim_{\delta \to 0} \frac{\int_{-\infty}^{x} \delta f(x, y)dx}{\delta f(y)} = \int_{-\infty}^{x} \frac{f(x, y)}{f(y)}dx.$$

Desta forma, definindo $P(X \leq x|Y = y)$ como a função de distribuição acumulada condicional de X dado $Y = y$, $F_{X|Y}(x|y)$, como uma densidade é a derivada da distribuição acumulada, então,

Definição 5.3.4: A *densidade condicional* de X dada $Y = y$ é:

$$f(x \mid y) = \begin{cases} \frac{f(x,y)}{f_Y(y)}, & (x, y) \in \mathbb{R}^2,\ y, \text{fixo, e } f_Y(y) > 0, \\ 0, & \text{quaisquer outros valores,} \end{cases}$$

A expressão acima é uma densidade pois:

(i) $f(x \mid y) \geq 0$, $\forall (x, y)$ porque é quociente de densidades.

(ii)

$$\begin{aligned}\int_{-\infty}^{+\infty} f(x \mid y)dx &= \int_{-\infty}^{+\infty} \frac{f(x,y)dx}{f_Y(y)} \\ &= \frac{1}{f_Y(y)}\int_{-\infty}^{+\infty} f(x,y)dx \\ &= \frac{f_Y(y)}{f_Y(y)} = 1.\end{aligned}$$

De forma similar,

$$f(y \mid x) = \begin{cases} \frac{f(x,y)}{f_X(x)}, & (x,y) \in \mathbb{R}^2,\ x, \text{fixo, e } f_X(x) > 0, \\ 0, & \text{quaisquer outros valores,} \end{cases}$$

Exemplo 5.3.5: A Tabela 5.1 contém a distribuição de probabilidade do sistema (X, Y), onde as variáveis aleatórias X e Y representam, respectivamente, o número de itens em *bom estado* e o número de itens em *perfeito estado*. Encontre as distribuições marginais e as condicionais.

Tabela 5.1: $P(X = j, Y = i)$.

$j \downarrow, i \rightarrow$	0	1	2	3	4	5	6
0	0.202	0.174	0.113	0.062	0.049	0.023	0.004
1	0	0.099	0.064	0.040	0.031	0.020	0.006
2	0	0	0.031	0.025	0.018	0.013	0.008
3	0	0	0	0.001	0.002	0.004	0.011

Solução: A Tabela 5.2 mostra as distribuições marginais, onde, usualmente, são exibidas.

Tabela 5.2: $P(X = j, Y = i)$ com marginais.

$j \downarrow, i \rightarrow$	0	1	2	3	4	5	6	$P(Y = i)$
0	0.202	0.174	0.113	0.062	0.049	0.023	0.004	0.627
1	0	0.099	0.064	0.040	0.031	0.020	0.006	0.260
2	0	0	0.031	0.025	0.018	0.013	0.008	0.095
3	0	0	0	0.001	0.002	0.004	0.011	0.018
$P(X = j)$	0.202	0.273	0.208	0.128	0.100	0.06	0.029	1

Algumas marginais:

$$P(X = 0 \mid Y = 0) = \frac{0.202}{0.202} = 1,$$

$$P(X = 1 \mid Y = 3) = \frac{0.040}{0.128} = 0.312,$$
$$P(Y = 6 \mid X = 2) = \frac{0.008}{0.208} = 0.038,$$
$$P(Y = 6 \mid X = 1) = \frac{0.006}{0.273} = 0.022.$$

■

Exemplo 5.3.6: Determine as densidades marginais $f_X(x)$ e $f_Y(y)$ e as densidades condicionais: de X dada Y, $f(x \mid y)$, e de Y dada X, $f(y \mid x)$ quando

$$f(x,y) = \begin{cases} x+y, & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{caso contrário.} \end{cases}$$

Solução: Obtendo as densidades marginais,

$$f_X(x) = \int_0^1 (x+y)dy = x + \frac{1}{2}, \text{ se } 0 \leq x \leq 1,$$
$$f_Y(y) = \int_0^1 (x+y)dx = y + \frac{1}{2}, \text{ se } 0 \leq y \leq 1.$$

Logo, as densidades condicionais são:

$$f(x|y) = \frac{x+y}{y+\frac{1}{2}}, \text{ se } 0 \leq x \leq 1, 0 \leq y \leq 1,$$
$$f(y|x) = \frac{x+y}{x+\frac{1}{2}}, \text{ se } 0 \leq x \leq 1, 0 \leq y \leq 1.$$

■

Exemplo 5.3.7: Determine as densidades marginais $f_X(x)$ e $f_Y(y)$ e as densidades condicionais: de X dada Y, $f(x \mid y)$, e de Y dada X, $f(y \mid x)$ quando

$$f(x,y) = \begin{cases} e^{-(x+y)}, & x \geq 0, y \geq 0, \\ 0, & \text{caso contrário.} \end{cases}$$

Solução: Obtendo as densidades marginais,

$$f_X(x) = \int_0^\infty e^{-(x+y)}dy = e^{-x}, \text{ se } x \geq 0,$$
$$f_Y(y) = \int_0^\infty e^{-(x+y)}dx = e^{-y}, \text{ se } y \geq 0.$$

Logo, as densidades condicionais são:

$$f(x|y) = e^{-x}, \text{ se } x \geq 0, y \geq 0,$$

$$f(y|x) = e^{-y}, \text{ se } x \geq 0, y \geq 0.$$

■

5.4 Independência entre Variáveis Aleatórias

Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$. Informalmente, as variáveis aleatórias X_i's são independentes se, e somente se, quaisquer eventos determinados por qualquer grupo de variáveis aleatórias distintas são independentes; por exemplo, $[X_1 < 5]$, $[X_2 > 9]$, e $[0 < X_5 \leq 3]$ são independentes. Formalmente,

Definição 5.4.1: Um conjunto de variáveis aleatórias $\{X_1, \ldots, X_n\}$ é mutuamente independente se, e somente se, para quaisquer eventos Borelianos $B_1, \ldots, B_n$,

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} P(X_i \in B_i).$$

O próximo teorema estabelece três critérios para provar que um conjunto de variáveis aleatórias é mutuamente independente.

Teorema 5.4.2: *As seguintes condições são necessárias e suficientes para testar se um conjunto $\{X_1, \ldots, X_n\}$ de variáveis aleatórias é mutuamente independente:*

(i) $F_{\vec{X}}(\vec{x}) = \prod_{i=1}^{n} F_{X_i}(x_i)$.

(ii) *Se $\vec{X}$ for um vetor aleatório discreto,*

$$p_{\vec{X}}(\vec{x}) = \prod_{i=1}^{n} p_{X_i}(x_i).$$

(iii) *Se $\vec{X}$ for um vetor aleatório contínuo,*

$$f_{\vec{X}}(\vec{x}) = \prod_{i=1}^{n} f_{X_i}(x_i), \forall (x_1, \ldots, x_n) \in I\!R^n.$$

Prova:

(i) Se $\{X_1, \ldots, X_n\}$ são variáveis aleatórias mutuamente independentes, então

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$
$$= \prod_{i=1}^{n} P(X_i \leq x_i) = \prod_{i=1}^{n} F_{X_i}(x_i), \forall (x_1, \ldots, x_n).$$

A prova da suficiência foge ao escopo deste livro (Burrill, (1972); Shiryayev, (1984)).

(ii) Se $\{X_1, \ldots, X_n\}$ são variáveis aleatórias mutuamente independentes, então

$$p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$
$$= \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} p_{X_i}(x_i), \forall (x_1, \ldots, x_n).$$

Reciprocamente, se a função de probabilidade de massa conjunta fatora e se $\{x_{i1}, x_{i2}, \ldots, x_{in}, \ldots\}$ são os possíveis valores assumidos pela variável aleatória X_i, então

$$\begin{aligned} P(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) &= \sum_{i:x_{1i}\in B_1} \cdots \sum_{i:x_{ni}\in B_n} P(X_1 = x_{1i}, \ldots, X_n = x_{ni}) \\ &= \sum_{i:x_{1i}\in B_1} \cdots \sum_{i:x_{ni}\in B_n} p_{X_1,\ldots,X_n}(x_{1i}, \ldots, x_{ni}) \\ &= \sum_{i:x_{1i}\in B_1} \cdots \sum_{i:x_{ni}\in B_n} \prod_{j=1}^{n} p_{X_j}(x_{ji}) \\ &= \prod_{j=1}^{n} P(X_j \in B_j). \end{aligned}$$

(iii) Consequência direta de (i) e da definição de função densidade.

∎

É fácil observar utilizando a definição de probabilidade condicional que se X e Y são independentes, então para todo A e B boreliano tal que $P(Y \in B) > 0$,

$$P(X \in A|Y \in B) = P(X \in A),$$

ou seja, se X e Y são independentes o conhecimento do valor de Y não altera a descrição probabilística de X.

Sobre como obter densidades conjuntas

(i) Encontrar a densidade conjunta do vetor (X, Y), por exemplo, pode ser não-trivial. A dificuldade inicial reside no fato de qual função $f(x, y)$ propor para densidade e como calcular seus parâmetros. Neste caso, métodos numéricos têm de ser usados.

(ii) Se as variáveis $X_1, \ldots, X_n$ forem independentes, então, como visto anteriormente, é fácil encontrar a densidade conjunta pois esta é o produto das marginais, isto é,

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \times \ldots \times f_{X_n}(x_n).$$

(iii) Se for factível supor que todos os pontos de uma dada região $R \subset \mathbb{R}^n$ estejam uniformemente dispostos em R, então a densidade conjunta é constante em R e sendo assim pode ser definida por

$$f(x_1, \ldots, x_n) = \begin{cases} \frac{1}{vol\ R}, & (x_1, \ldots, x_n) \in R, \\ 0, & \text{caso contrário.} \end{cases}$$

Se $n = 1$, $vol\ R$ é o comprimento do intervalo, isto é, neste caso $R = [a, b]$, $vol\ R = b - a$ e $f_X(x) = \frac{1}{b-a}$ que é a uniforme no intervalo $[a, b]$. Se $n = 2$, $vol\ R =$ área de R.

Exemplo 5.4.3: Uma variável aleatória contínua tem função densidade conjunta $f(x, y) = 15x^2y$ definida no triângulo (0,0), (1,0) e (0,2). Determine as densidades marginais e verifique se X e Y são independentes.
Solução: Obtendo as densidades marginais,

$$f_X(x) = \int_0^{2-2x} 15x^2y dy = 30x^2(x-1)^2, \text{ se } 0 \le x \le 1,$$

$$f_Y(y) = \int_0^{\frac{2-y}{2}} 15x^2y dx = \frac{5y(2-y)^3}{8}, \text{ se } 0 \le y \le 2.$$

Como $f(x, y) \neq f_X(x)f_Y(y)$, as variáveis aleatórias não são independentes. ■

Exemplo 5.4.4: Suponha a seguinte distribuição de probabilidade conjunta do vetor (X, Y) onde somente as probabilidades abaixo são não-nulas,

$$P(X = 1, Y = 2) = P(X = 1, Y = 3) = P(X = 1, Y = 4) = 1/6,$$

$$P(X = 2, Y = 3) = P(X = 2, Y = 4) = 1/6,$$

$$P(X = 3, Y = 4) = 1/6.$$

(a) Estabeleça a distribuição de probabilidade da variável X.

Solução: $P(X = 1) = \frac{3}{6}$, $P(X = 2) = \frac{2}{6}$, $P(X = 3) = \frac{1}{6}$, $P(X = k) = 0, \forall k \neq 1, 2, 3$.

(b) Verifique se X e Y são independentes.

Solução: $P(X = 2) = \frac{2}{6}$, $P(Y = 2) = \frac{1}{6}$, então, $\exists (i, j) = (2, 2)$ pertencente ao contradomínio de (X, Y) tal que $P(X = i, Y = j) \neq P(X = i)P(Y = j)$. Portanto X e Y não são independentes. ■

Exemplo 5.4.5: Considere uma amostra aleatória de tamanho 2 tirada de um conjunto de 10 aplicações entre as quais tem-se 2 *telnet*, *1 ftp*, *2 teleconferência*, *3 web* e *2 correio eletrônico*. Definindo as variáveis aleatórias X_1 e X_2 como se segue: para $k = 1, 2$, $X_k = 1$ ou $X_k = 0$, se, respectivamente, a k-ésima aplicação é *web*, ou não. Descrever a lei de probabilidade conjunta do vetor (X_1, X_2) e as distribuições marginais se a amostra foi retirada

(a) sem reposição;

Solução:

$$P(X_1 = i, X_2 = j)$$

$i \downarrow, j \rightarrow$	0	1	$P(X_2 = j)$
0	$\frac{42}{90}$	$\frac{21}{90}$	$\frac{63}{90}$
1	$\frac{21}{90}$	$\frac{6}{90}$	$\frac{27}{90}$
$P(X_1 = i)$	$\frac{63}{90}$	$\frac{27}{90}$	1

$$P(X_1 = 0, X_2 = 0) = \frac{7}{10} \times \frac{6}{9} = \frac{42}{90}.$$
$$P(X_1 = 0, X_2 = 1) = \frac{7}{10} \times \frac{3}{9} = \frac{21}{90}.$$
$$P(X_1 = 1, X_2 = 0) = \frac{3}{10} \times \frac{7}{9} = \frac{21}{90}.$$
$$P(X_1 = 1, X_2 = 1) = \frac{3}{10} \times \frac{2}{9} = \frac{6}{90}.$$

(b) com reposição.

Solução:

$$P(X_1 = i, X_2 = j)$$

$i \downarrow, j \rightarrow$	0	1	$P(X_2 = j)$
0	$\frac{49}{100}$	$\frac{21}{100}$	$\frac{70}{100}$
1	$\frac{21}{100}$	$\frac{9}{100}$	$\frac{30}{100}$
$P(X_1 = i)$	$\frac{70}{100}$	$\frac{30}{100}$	1

$$P(X_1 = 0, X_2 = 0) = \frac{7}{10} \times \frac{7}{10} = \frac{49}{100}.$$
$$P(X_1 = 0, X_2 = 1) = \frac{7}{10} \times \frac{3}{10} = \frac{21}{100}.$$
$$P(X_1 = 1, X_2 = 0) = \frac{3}{10} \times \frac{7}{10} = \frac{21}{100}.$$
$$P(X_1 = 1, X_2 = 1) = \frac{3}{10} \times \frac{3}{10} = \frac{9}{100}.$$

■

Exemplo 5.4.6: Um programa é composto de dois módulos cujos tempos de execução X e Y são variáveis aleatórias independentes uniformemente distribuídas sobre $\{1, 2, \cdots, n\}$.

(a) Encontre a distribuição conjunta do vetor (X, Y).

Solução:

$$P(X = i, Y = j) = \frac{1}{n^2}, \forall i, j = 1, 2, \ldots, n.$$

(b) Calcule $P(X = Y)$, $P(X \leq Y)$ e $P(X + Y \geq 3)$.

Solução:

$$P(X = Y) = P(X = 1, Y = 1) + \cdots + P(X = n, Y = n) = n \times \frac{1}{n^2} = \frac{1}{n}.$$

$$P(X \leq Y) = \frac{n}{n^2} + \frac{n-1}{n^2} + \cdots + \frac{1}{n^2} = \frac{n+1}{2n}.$$

$$P(X + Y \geq 3) = 1 - P(X + Y < 3) = 1 - P(X = 1, Y = 1) = 1 - \frac{1}{n^2} = \frac{n^2 - 1}{n^2}.$$

■

Exemplo 5.4.7: Sejam X e Y variáveis aleatórias discretas independentes tais que $P(X = k) = P(Y = K) = 2^{-k}$, para $k = 1, 2, \cdots,$. Calcule

(a) $P(X = Y)$.

Solução:

$$P(X = Y) = P(X = 1, Y = 1) + P(X = 2, Y = 2) + \cdots = 2^{-2} + 2^{-4} + \cdots = \frac{1}{3}.$$

(b) $P(X > Y)$.

Solução: $P(X > Y) + P(X < Y) + P(X = Y) = 1$. Mas, $P(X > Y) = P(X < Y)$. Logo,

$$2P(X > Y) + P(X = Y) = 2P(X > Y) + \frac{1}{3} = 1 \Rightarrow P(X > Y) = \frac{1}{3}.$$

■

Exemplo 5.4.8: Sejam X e Y variáveis aleatórias tendo distribuição de probabilidade conjunta dada por

$X \downarrow, Y \rightarrow$	-1	0	2	6
-2	$1/9$	$1/27$	$1/27$	$1/9$
1	$2/9$	0	$1/9$	$1/9$
3	0	0	$1/9$	$4/27$

Determine

(a) a probabilidade de $XY = 2$;

Solução:

$$P(XY = 2) = P(X = -2, Y = -1) + P(X = 1, Y = 2) = 2/9.$$

(b) as distribuições marginais;

Solução:

$X \downarrow, Y \rightarrow$	-1	0	2	6	$P(X = i)$
-2	$1/9$	$1/27$	$1/27$	$1/9$	$8/27$
1	$2/9$	0	$1/9$	$1/9$	$4/9$
3	0	0	$1/9$	$4/27$	$7/27$
$P(Y = j)$	$3/9$	$1/27$	$7/27$	$10/27$	1

(c) X e Y são independentes?

Solução: Não porque $P(X = 3, Y = -1) = 0 \neq P(X = 3)(P(Y = -1)$. ■

Exemplo 5.4.9: Seja o vetor aleatório (X, Y) uniformemente distribuído na região $R_{(X,Y)} = \{(x, y) \mid x > 0, y > 0,\ x + y < 1\}$. Calcule

(a) a conjunta $f(x, y)$;

Solução:

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

(b) a densidade da variável aleatória X;

Solução:

$$f_X(x) = \int_0^{-x+1} 2dy = 2(1 - x), 0 < x < 1.$$

(c) a densidade da variável aleatória Y;

Solução:

$$f_Y(y) = \int_0^{-y+1} 2dx = 2(1 - y), 0 < y < 1.$$

(d) Verifique se X e Y são independentes.

Solução:

$$f_X(x)f_Y(y) = 2(1 - x)2(1 - y) \neq 2 = f(x, y).$$

X e Y não são independentes. ■

Exemplo 5.4.10: Considere a variável aleatória bidimensional (X, Y) uniformemente distribuída na região poligonal T de vértices (-2,0), (2,0), (1,1) e (-1,1).

(a) Determine a função de densidade de probabilidade conjunta $f(x, y)$.

Solução:

$$f(x,y) = \begin{cases} \frac{1}{3}, & (x,y) \in T, \\ 0, & \text{caso contrário.} \end{cases}$$

(b) Determine a função de densidade de probabilidade marginal $f_X(x)$.

Solução:

$$f_X(x) = \begin{cases} \int_0^1 \frac{1}{3} dy = \frac{1}{3}, & 0 \leq \mid x \mid \leq 1, \\ \int_0^{2-|x|} \frac{1}{3} dy = \frac{2-|x|}{3}, & 1 \leq \mid x \mid \leq 2, \\ 0, & \text{caso contrário.} \end{cases}$$

(c) Determine a função de densidade de probabilidade marginal $f_Y(y)$.

Solução:

$$f_Y(y) = \begin{cases} \int_{-(2-y)}^{2-y} \frac{1}{3} dx = \frac{2}{3}(2-y), & 0 < y < 1, \\ 0, & \text{caso contrário.} \end{cases}$$

(d) Verifique se X e Y são variáveis aleatórias independentes.

Solução: $f(x,y) = \frac{1}{3} \neq f_X(x) f_Y(y)$. Portanto, X e Y não são independentes. ■

Exemplo 5.4.11: Seja

$$f(x,y) = \begin{cases} 2, & 0 < x < y,\ 0 < y < 1, \\ 0, & \text{quaisquer outros casos,} \end{cases}$$

a função densidade conjunta do vetor aleatório (X, Y).

(a) Calcule as densidades das variáveis aleatórias X e Y.

Solução:

$$f_X(x) = \int_x^1 f(x,y) dy = 2(1-x), 0 < x < 1.$$

$$f_Y(y) = \int_0^y f(x,y) dx = 2y, 0 < y < 1.$$

(b) X e Y são independentes? Justifique sua resposta.

Solução: $f_X(x) f_Y(y) = 2(1-x)2y \neq 2 = f(x,y)$. Não são independentes. ■

5.5 Funções de Vetores Aleatórios

O objetivo nesta seção é, considerando o vetor aleatório (X,Y) onde X e Y são variáveis aleatórias definidas no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$, encontrar a distribuição de probabilidade de $Z = H(X,Y)$ sendo H uma função real tal que seu domínio contém os contradomínios de X e Y, respectivamente, R_X e R_Y.

Quando necessário, os resultados serão mostrados para vetores n-dimensionais, quando não, para vetores bidimensionais. Já é um bom começo entender o procedimento para $n = 2$.

Considere primeiro o caso em que $\vec{X}$ é um vetor aleatório discreto. Se $\vec{Y} = H(\vec{X})$ e sendo $\vec{x}_{i1}, \vec{x}_{i2}, \vec{x}_{i3}, \ldots$ os valores de $\vec{X}$ tal que $H(\vec{x}_{ij}) = \vec{y}_i$ para todo j. Então,

$$P(\vec{Y} = \vec{y}_i) = P(\vec{X} \in \{\vec{x}_{i1}, \vec{x}_{i2}, \vec{x}_{i3}, \ldots\}) = \sum_{j=1}^{\infty} P(\vec{X} = \vec{x}_{ij}) = \sum_{j=1}^{\infty} p_{\vec{X}}(\vec{x}_{ij}),$$

ou seja, para calcular a probabilidade do evento $\{\vec{Y} = \vec{y}_i\}$, acha-se o evento equivalente em termos de $\vec{X}$, isto é, todos os valores $\vec{x}_{ij}$ de $\vec{X}$ tal que $H(\vec{x}_{ij}) = \vec{y}_i$ e somam-se as probabilidades de $\vec{X}$ assumir cada um desses valores.

Seja agora o caso em que (X,Y) e $Z = H(X,Y)$ são contínuos. Fixado z, a solução geral do problema é:

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P(H(X,Y) \leq z) \\ &= P((X,Y) \in B_z) \\ &= \int \int_{B_z} f(x,y)dxdy, \end{aligned}$$

onde $B_z \subseteq \mathbb{R}^2$, $B_z \in \mathcal{B}^2$, isto é, B_z é um elemento da σ-álgebra de Borel sobre $\mathbb{R}^2$,

$$B_z = \{(x,y) : H(x,y) \leq z\}.$$

Se for possível obter uma função $g \geq 0$ tal que

$$\int \int_{B_z} f(x,y)dxdy = \int_{-\infty}^{z} g(v)dv$$

então,

$$g(\cdot) = f_Z(\cdot),$$

isto é, g é a densidade de Z, $f_Z(\cdot)$.

O que será feito a seguir é como usar este resultado para encontrar a distribuição da soma, produto e quociente de X e Y.

5.5.1 Distribuição de $Z = X + Y$

Seja $Z = X + Y$ e z fixo. Então,

$$\begin{aligned} B_z &= \{(x, y) : x + y \leq z\} \\ &= \{(x, y) : -\infty < x < +\infty, -\infty < y \leq z - x\}. \end{aligned}$$

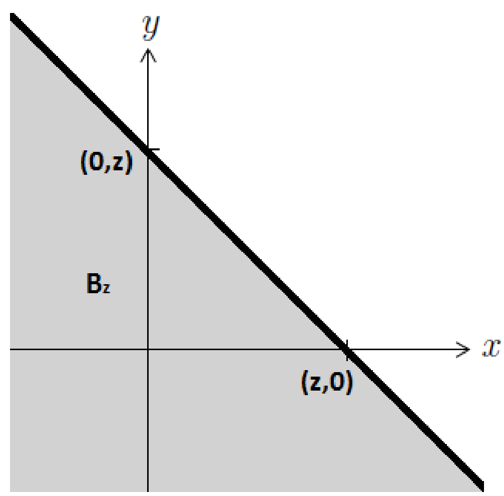

Figura 5.1: Região B_z

$$\begin{aligned} F_Z(z) &= \int\int_{B_z} f(x, y)dxdy \\ &= \int_{-\infty}^{+\infty}(\int_{-\infty}^{z-x} f(x, y)dy)dx. \end{aligned}$$

Fazendo uma mudança de variável na integral interna:

$$y = v - x \Rightarrow dy = dv.$$

Como $y \leq z - x$ então $v - x \leq z - x \Rightarrow v \leq z$. Logo, $-\infty < v \leq z < +\infty$ e portanto v varia de $-\infty$ a z. Assim.

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^{+\infty}(\int_{-\infty}^{z} f(x, v - x)dv)dx \\ &= \int_{-\infty}^{z}(\int_{-\infty}^{+\infty} f(x, v - x)dx)dv. \end{aligned}$$

Logo, $g(v) = \int_{-\infty}^{+\infty} f(x, v - x)dx$ e

$$f_Z(z) = f_{X+Y}(z) = \int_{-\infty}^{+\infty} f(x, z-x)dx, -\infty < z < +\infty. \tag{5.7}$$

Se X e Y forem independentes (5.7) torna-se

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx, -\infty < z < +\infty. \tag{5.8}$$

De (5.8) tem-se que a densidade da soma de duas variáveis aleatórias independentes é a *convolução* das densidades marginais.

Se X e Y forem independentes e não-negativas (5.7) torna-se

$$f_{X+Y}(z) = \int_0^z f_X(x)f_Y(z-x)dx, z > 0.$$

Exemplo 5.5.1: Suponha que X e Y têm densidade valendo 1 no intervalo [0,1] e que são independentes. Encontrar a densidade de $S = X + Y$.
Solução: Do problema sabe-se que

$$f_X(x) = 1, 0 \le x \le 1$$

e

$$f_Y(y) = 1, 0 \le y \le 1.$$

Seja $S = X + Y$. Logo,

$$f_S(s) = \int f_X(x)f_Y(s-x)dx, 0 \le x \le 1,\ 0 \le y \le 1.$$

Como $0 \le x \le 1$ e $0 \le y \le 1$ então

$$0 \le x \le 1 \wedge 0 \le s-x \le 1 \Rightarrow 0 \le x \le 1 \wedge s-1 \le x \le s.$$

A Figura 5.2 ilustra as situações possíveis para s.

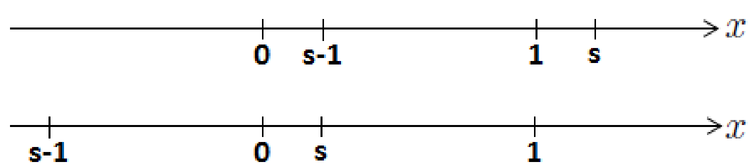

Figura 5.2: Situações possíveis para s

(a) $s-1 \le 0 \wedge 0 \le s \le 1 \Rightarrow 0 \le s \le 1.$

(b) $0 \le s-1 \le 1 \wedge s \ge 1 \Rightarrow 0 < s \le 2 \wedge s \ge 1 \Rightarrow 1 \le s \le 2.$

Se $0 \leq s \leq 1$ então $0 \leq x \leq s$; se $1 \leq s \leq 2$, então $s-1 \leq x \leq 1$. Logo,

$$f_S(s) = \int_0^s dx = s, 0 \leq s \leq 1,$$

$$f_S(s) = \int_{s-1}^1 dx = 2-s, 1 \leq s \leq 2.$$

Concluindo,

$$f_S(s) = \begin{cases} s, & 0 \leq s \leq 1, \\ 2-s, & 1 \leq s \leq 2, \\ 0, & \text{quaisquer outros valores.} \end{cases}$$

■

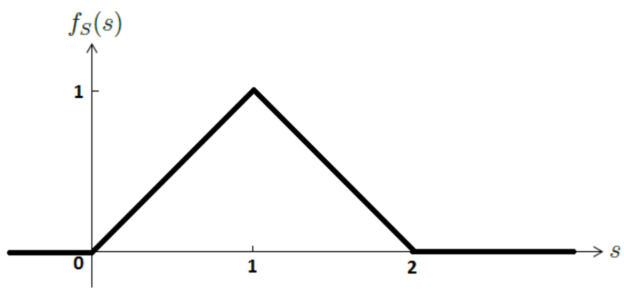

Figura 5.3: Convolução de 2 uniformes

Exemplo 5.5.2: Sejam X e Y com densidade conjunta dada por

$$f(x, y) = e^{-(x+y)}, x \geq 0, y \geq 0.$$

Encontre a densidade de $V = X + Y$.
Solução:

$$f_V(v) = \int_{-\infty}^{+\infty} f(x, v-x)dx.$$

Mas, $v - x \geq 0 \Rightarrow x \leq v$ e $e^{-(x+y)} = e^{-x}e^{-(v-x)} = e^{-v}$. Logo

$$f_V(v) = \int_0^v e^{-v}dx = ve^{-v}, v \geq 0.$$

■

Exemplo 5.5.3: Se as variáveis aleatórias X_1 e X_2 são independentes e identicamente distribuídas com a densidade

$$f(t) = \begin{cases} te^{-t}, & t \geq 0, \\ 0, t < 0, \end{cases}$$

encontre a densidade de $S = X_1 + X_2$.

Solução:

$$f_{X_1}(x_1) = x_1 e^{-x_1}, x_1 \geq 0,$$

$$f_{X_2}(x_2) = x_2 e^{-x_2}, x_2 \geq 0.$$

Logo,

$$f_S(s) = \int_{-\infty}^{+\infty} f_{X_1}(x_1) f_{X_2}(s - x_1) dx_1, s \geq 0.$$

Mas, $x_1 \geq 0$ e $s - x_1 \geq 0 \Rightarrow s \geq x_1$. Logo, $0 \leq x_1 \leq s$. Portanto,

$$\begin{aligned} f_S(s) &= \int_0^s f_{X_1}(x_1) f_{X_2}(s - x_1) dx_1 \\ &= \int_0^s x_1 e^{x_1} (s - x_1) e^{-(s - x_1)} dx_1 \\ &= \frac{e^{-s} s^3}{6}. \end{aligned}$$

Como $s \geq x_1$ e $x_1 \geq 0$ então $s \geq 0$. Assim,

$$f_S(s) = \frac{e^{-s} s^3}{6}, s \geq 0.$$

■

Exemplo 5.5.4: Sejam X e Y variáveis aleatória independentes e identicamente distribuídas com densidade $f(t) = e^{-t}, t \geq 0$. Encontre a densidade de $S = X + Y$.
Solução:

$$f_X(x) = e^{-x}, x \geq 0,$$

$$f_Y(y) = e^{-y}, y \geq 0,$$

logo,

$$f_S(s) = \int_0^{+\infty} f_X(x) f_Y(s - x) dx, s \geq 0.$$

Entretanto, de $x \geq 0$ e $s - x \geq 0$ o que implica que $s \geq x$ tem-se que $0 \leq x \leq s$. Portanto,

$$f_S(s) = \int_0^s e^{-x} e^{-(s-x)} dx = s e^{-s}.$$

Logo,

$$f_S(s) = s e^{-s}, s \geq 0.$$

■

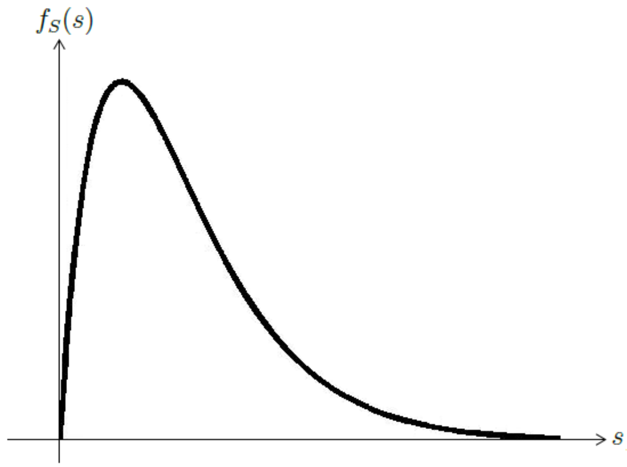

Figura 5.4: Gráfico de $f_S(s)$

5.5.2 Distribuição de $Z = XY$

Seja $Z = XY$, isto é, $H(X,Y) = XY$. Fixando z, então

$$B_z = \{(x,y) : xy \leq z\}.$$

Se

$$x > 0,\ xy \leq z \Rightarrow y \leq \frac{z}{x},$$
$$x < 0,\ xy \leq z \Rightarrow y \geq \frac{z}{x}.$$

Logo,

$$B_z = \{(x,y) : -\infty < x < 0 \wedge y \geq \frac{z}{x}\} \cup \{(x,y) : 0 < x < +\infty \wedge y \leq \frac{z}{x}\} = B_1 \cup B_2.$$

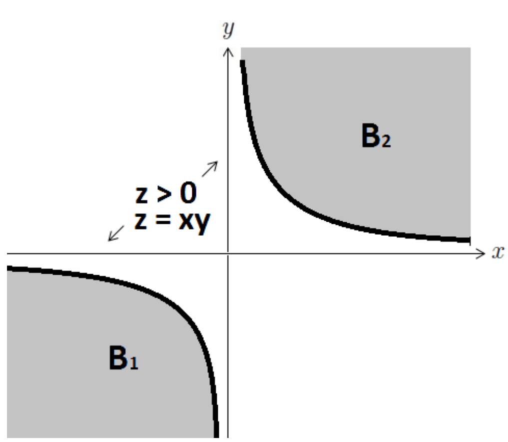

Figura 5.5: Regiões B_1 e B_2 para $z = xy$, onde $z > 0$

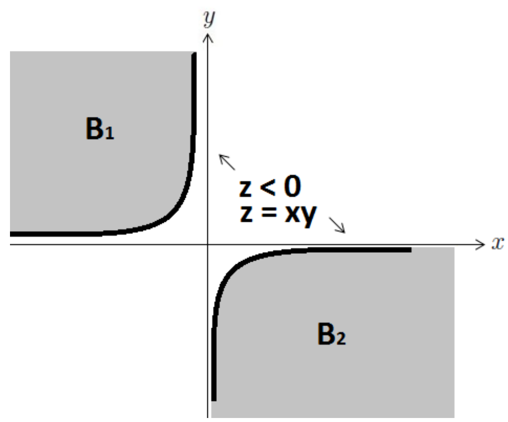

Figura 5.6: Regiões B_1 e B_2 para $z = xy$, onde $z < 0$

Então

$$\begin{aligned} F_Z(z) &= \int\int_{B_z} f(x,y)dxdy \\ &= \int_{-\infty}^{0}(\int_{\frac{z}{x}}^{+\infty} f(x,y)dy)dx + \int_{0}^{+\infty}(\int_{-\infty}^{\frac{z}{x}} f(x,y)dy)dx. \end{aligned}$$

Fazendo uma mudança de variável na integral interna:

$$y = \frac{v}{x} \Rightarrow dy = \frac{1}{x}dv.$$

Substituindo o valor de y em B_1 e B_2,

$$\frac{v}{x} \geq \frac{z}{x} \Rightarrow v \geq z \Rightarrow z \leq v < +\infty,$$

$$\frac{v}{x} \leq \frac{z}{x} \Rightarrow v \leq z \Rightarrow -\infty < v \leq z.$$

Logo,

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^{0}(\int_{z}^{-\infty} f(x,\frac{v}{x})\frac{1}{x}dv)dx + \int_{0}^{+\infty}(\int_{-\infty}^{z} f(x,\frac{v}{x})\frac{1}{x}dv)dx \\ &= \int_{-\infty}^{0}(\int_{-\infty}^{z} (-\frac{1}{x})f(x,\frac{v}{x})dv)dx + \int_{-\infty}^{0}(\int_{-\infty}^{z} \frac{1}{x}f(x,\frac{v}{x})dv)dx \\ &= \int_{-\infty}^{+\infty}(\int_{-\infty}^{z} \mid \frac{1}{x} \mid f(x,\frac{v}{x})dx)dv \\ &= \int_{-\infty}^{z}(\int_{-\infty}^{+\infty} \mid \frac{1}{x} \mid f(x,\frac{v}{x})dx)dv. \end{aligned}$$

Portanto,

$$f_{XY}(z) = \int_{-\infty}^{+\infty} \mid \frac{1}{x} \mid f(x, \frac{z}{x})dx, -\infty < z < +\infty. \tag{5.9}$$

Se X e Y forem independentes, de (5.9) tem-se

$$f_{XY}(z) = \int_{-\infty}^{+\infty} \mid \frac{1}{x} \mid f_X(x)f_Y(\frac{z}{x})dx, -\infty < z < +\infty.$$

Se X e Y forem independentes e não-negativas,

$$f_{XY}(z) = \int_{0}^{+\infty} \frac{1}{x} f_X(x)f_Y(\frac{z}{x})dx, z > 0.$$

Exemplo 5.5.5: Seja $f(x,y) = 1, 0 \leq x \leq 1, 0 \leq y \leq 1$. Determinar a densidade de $Z = XY$.

Solução: Como $z = xy$ então $y = \frac{z}{x}$ ou $x = \frac{z}{y}$. Adicionalmente,

$$0 \leq x \leq 1 \Rightarrow 0 \leq x = \frac{z}{y} \leq 1,$$

$$0 \leq y \leq 1 \Rightarrow 0 \leq y = \frac{z}{x} \leq 1$$

e portanto, $z \leq x \leq 1$. Logo,

$$f_Z(z) = \int_{z}^{1} \frac{1}{x} f(x, \frac{z}{x})dx = -lnz, 0 < z < 1.$$

■

5.5.3 Distribuição de $Z = \frac{Y}{X}$

Seja $Z = \frac{Y}{X}$ e z fixo. Logo

$$B_z = \{(x,y) : \frac{y}{x} \leq z\}.$$

Se,

$$x > 0, \frac{y}{x} \leq z \Rightarrow y \leq xz,$$
$$x < 0, \frac{y}{x} \leq z \Rightarrow y \geq xz.$$

Portanto,

$$\begin{aligned} B_z &= \{(x,y) : -\infty < x < 0 \wedge y \geq xz\} \cup \{(x,y) : 0 < x < +\infty \wedge y \leq xz\} \\ &= B_1 \cup B_2. \end{aligned}$$

Então,

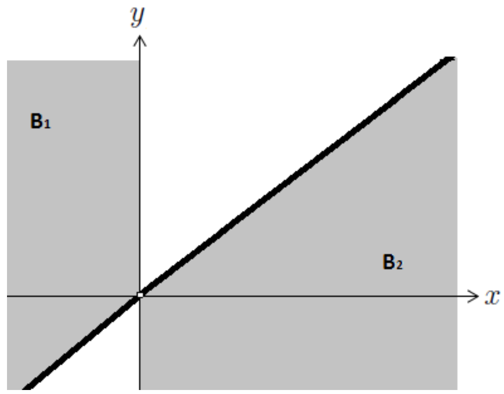

Figura 5.7: Regiões B_1 e B_2 para $z = \frac{y}{x}$, onde $z > 0$

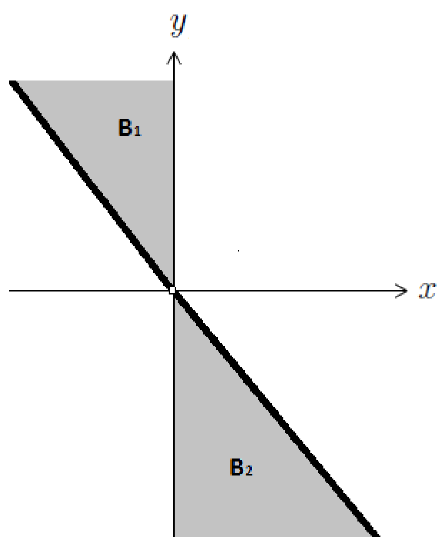

Figura 5.8: Regiões B_1 e B_2 para $z = \frac{y}{x}$, onde $z < 0$

$$\begin{aligned} F_z(z) &= \int\int_{B_z} f(x,y)dxdy \\ &= \int_{-\infty}^{0}(\int_{xz}^{+\infty} f(x,y)dy)dx + \int_{0}^{+\infty}(\int_{-\infty}^{xz} f(x,y)dy)dx. \end{aligned}$$

Fazendo uma mudança de variáveis na integral mais interna e substituindo no valor de y em B_z tem-se:

$$y = xv \Rightarrow dy = xdv$$

e

$$xv \geq xz \Rightarrow v \geq z \Rightarrow z \leq v < +\infty,$$

$$xv \leq xz \Rightarrow v \leq z \Rightarrow -\infty < v \leq z.$$

Assim,

$$\begin{aligned}
F_Z(z) &= \int_{-\infty}^{0}(\int_{z}^{-\infty} xf(x,xv)dv)dx + \int_{0}^{+\infty}(\int_{-\infty}^{z} xf(x,xv)dv)dx \\
&= \int_{-\infty}^{0}(\int_{-\infty}^{z} (-x)f(x,xv)dv)dx + \int_{0}^{+\infty}(\int_{-\infty}^{z} xf(x,xv)dv)dx \\
&= \int_{-\infty}^{+\infty}(\int_{-\infty}^{z} \mid x \mid f(x,xv)dv)dx \\
&= \int_{-\infty}^{z}(\int_{-\infty}^{+\infty} \mid x \mid f(x,xv)dx)dv
\end{aligned}$$

Logo,

$$f_{\frac{Y}{X}}(z) = \int_{-\infty}^{+\infty} \mid x \mid f(x,xz)dx, -\infty < z < +\infty.$$

Se X e Y forem independentes,

$$f_{\frac{Y}{X}}(z) = \int_{-\infty}^{+\infty} \mid x \mid f_X(x)f_Y(xz)dx, -\infty < z < +\infty.$$

Se X e Y forem independentes e não-negativas,

$$f_{\frac{Y}{X}}(z) = \int_{0}^{+\infty} xf_X(x)f_Y(xz)dx, z > 0.$$

Exemplo 5.5.6: Sejam X e Y com densidade conjunta dada por

$$f(x,y) = e^{-(x+y)}, x \geq 0, y \geq 0.$$

Encontre a densidade de $U = X/Y$.
Solução: $U = X/Y \Rightarrow X = UY$. Logo, $x = uy \Rightarrow dx = ydu$. Também, $e^{-(x+y)} = e^{-x}e^{-y} = e^{-uy}e^{-y} = e^{-y(1+u)}$. Portanto,

$$f_{X/Y}(u) = \int_{-\infty}^{+\infty} \mid y \mid f(uy,y)dy = \int_{0}^{+\infty} ye^{-y(1+u)}dy = \frac{1}{(1+u)^2}, u \geq 0.$$

■

Exemplo 5.5.7: Duas pessoas marcam um encontro em determinado lugar entre 12:00 e 13:00. Cada uma chega, na hora marcada, ao encontro independentemente e com uma densidade constante. Ficou acertado entre ambas que nenhuma delas esperará mais do que 15 minutos pela outra. Determinar a probabilidade de se encontrarem.
Solução: Este problema será resolvido de três formas distintas. A figura a seguir ilustra a região de encontro, E, de ambas.

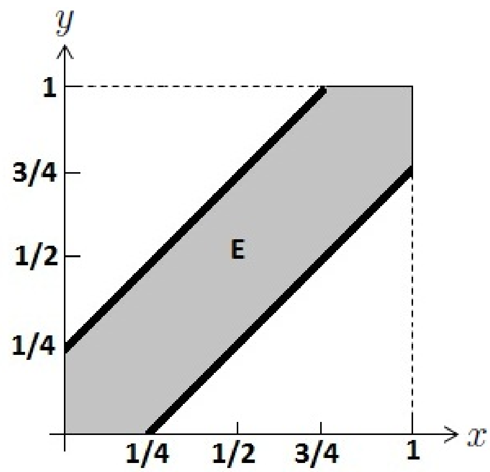

Figura 5.9: Região de encontro E

1. Usando probabilidade geométrica.

   O quadrado de vértices (0,0), (0,1), (1,0) e (1,1) tem lado 1, consequentemente área $S = 1$. A região do encontro tem área $1 - (\frac{3}{4})^2 = \frac{7}{16}$. Logo, a probabilidade de que ambas se encontrem é $\frac{7}{16}$.

2. Usando densidade conjunta.

   Sejam X e Y, respectivamente, os tempos de chegadas das duas pessoas. De acordo com os dados do problema, como entre 12:00 e 13:00 tem-se uma hora,

   $$X \sim U(0,1)$$

   e

   $$Y \sim U(0,1).$$

   Como X e Y são independentes,

   $$f(x,y) = f_X(x)f_Y(y) = \begin{cases} 1, & 0 < x < 1, 0 < y < 1, \\ 0, & \text{quaisquer outros valores.} \end{cases}$$

   A probabilidade de se encontrarem em E é dada por:

   $$\begin{aligned} P(E) &= \int\int_E f(x,y)dxdy \\ &= \int\int_{R_1} f(x,y)dxdy + \int\int_{R_2} f(x,y)dxdy + \int\int_{R_3} f(x,y)dxdy. \end{aligned}$$

   Portanto,

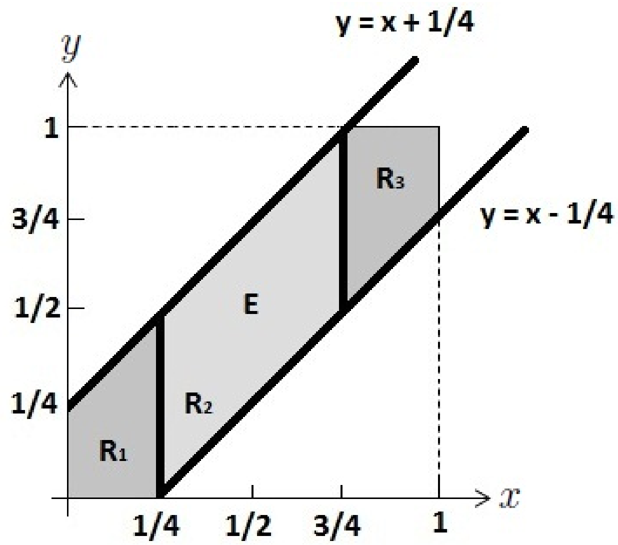

Figura 5.10: Regiões R_1, R_2 e R_3

$$\int\int_{R_1} f(x,y)dxdy = \int_0^{\frac{1}{4}} (\int_0^{x+\frac{1}{4}} dy)dx = \frac{3}{32},$$
$$\int\int_{R_2} f(x,y)dxdy = \int_{\frac{1}{4}}^{\frac{3}{4}} (\int_{x-\frac{1}{4}}^{x+\frac{1}{4}} dy)dx = \frac{4}{16},$$
$$\int\int_{R_1} f(x,y)dxdy = \int_{\frac{3}{4}}^{1} (\int_{x-\frac{1}{4}}^{1} dy)dx = \frac{3}{32}.$$

Logo,

$$P(E) = \frac{3}{32} + \frac{4}{16} + \frac{3}{32} = \frac{7}{16}.$$

3. Usando função de vetor aleatório.

Como visto anteriormente no exemplo (5.5.1), a densidade de $S = X + Y$, quando $X \sim U(0,1)$ e $Y \sim U(0,1)$ é $f_S(s) = \int_{-\infty}^{+\infty} f_X(x) f_Y(s-x)dx$. O problema proposto consiste em calcular

$$P(\mid X - Y \mid \leq \frac{1}{4}),$$

assim, a distribuição de interesse é em

$$Z = X - Y.$$

Por simetria, é fácil supor que

$$f_Z(z) = \int f_X(x) f_Y(x-z)dx$$

pois $z = x - y \Rightarrow y = x - z$.

De acordo com os dados do problema, o integrando será não nulo quando

$$0 \leq x \leq 1 \wedge 0 \leq x - z \leq 1. \tag{5.10}$$

De (5.10) tem-se que

$$0 \leq x \leq 1 \wedge z \leq x \leq z + 1 \tag{5.11}$$

A partir de (5.11) tem-se as seguintes situações:

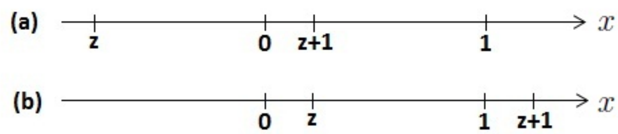

Figura 5.11: Situações possíveis para a variação de x

(a) $z \leq 0 \wedge 0 \leq z + 1 \leq 1 \Rightarrow z \leq 0 \wedge -1 \leq z \leq 0 \Rightarrow -1 \leq z \leq 0.$
(b) $0 \leq z \leq 1 \wedge \leq z + 1 \Rightarrow 0 \leq z \leq 1 \wedge z \geq 0 \Rightarrow 0 \leq z \leq 1.$

Em (a) x toma valores entre 0 e $z+1$; em (b) x varia de z a 1. Logo,

$$f_Z(z) = \int_0^{z+1} 1dx = 1 + z, -1 \leq z \leq 0,$$
$$f_Z(z) = \int_z^1 1dx = 1 - z, 0 \leq z \leq 1.$$

Portanto,

$$f_Z(z) = \begin{cases} 1+z, & -1 \leq z \leq 0, \\ 1-z, & 0 < z \leq 1, \\ 0, & \text{quaisquer outros valores.} \end{cases}$$

É fácil ver $\int_{-1}^1 f_Z(z)dz = 1$. A probabilidade pedida é:

$$\begin{aligned} P(\mid X - Y \mid \leq \frac{1}{4}) &= P(-\frac{1}{4} \leq Z \leq \frac{1}{4}) \\ &= \int_{-\frac{1}{4}}^{0} (1+z)dz + \int_0^{\frac{1}{4}} (1-z)dz \\ &= \frac{7}{16}. \end{aligned}$$

■

5.5.4 Jacobiano de uma Função

Os resultados vistos anteriormente sobre a distribuição da soma, produto e quociente de variáveis aleatórias também poderiam ter sido obtidos via Jacobiano de uma função, como a seguir.

Dado um conjunto de n equações em n variáveis $x_1, \ldots, x_n$,

$$y_1 = f_1(x_1, ..., x_n), \ldots, y_n = f_n(x_1, ..., x_n),$$

a matriz Jacobiana é definida por

$$J = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_n} \end{pmatrix}$$

O determinante de J é chamado de *Jacobiano*. Pode-se provar que o módulo do Jacobiano dá a razão entre volumes n-dimensionais em $\vec{y}$ e $\vec{x}$ quando a maior dimensão Δx_i tende a zero. Deste modo, o módulo do Jacobiano aparece nas mudanças de variáves de integração em integrais múltiplas, ou seja, existe um teorema do cálculo (Ávila, (1994); Guidorizzi, (1994)) que afirma que se $f : G_0 \to G$ for uma bijeção entre G_0 e G, f e as derivadas parcias que aparecem na matriz Jacobiana forem funções contínuas em G_0, e o Jacobiano for diferente de zero para todo $x \in G_0$, então

$$\int \cdots \int_A g(y_1, \ldots, y_n) dy_1 \cdots dy_n = \int \cdots \int_{f^{-1}(A)} g(f_1(x_1, ..., x_n), \ldots, f_n(x_1, ..., x_n))|J| dx_1 \cdots dx_n,$$

para qualquer função g integrável em $A \subseteq G$.

O conceito de Jacobiano será usado para resolver o seguinte exemplo da soma de duas variáveis aleatórias.

Exemplo 5.5.8: Suponha que (X, Y) tenha densidade conjunta $f(x, y)$ e seja $Z = X + Y$. Neste caso,

$$F_Z(z) = P(Z \leq z) = P(X + Y \leq z) = P((X, Y) \in B_z),$$

onde $B_z = \{(x, y) : x + y \leq z\}$. Portanto,

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f(x, y) dx dy.$$

Fazendo a mudança de variáveis $s = x + y$, $t = y$, que tem jacobiano igual a 1, tem-se

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z} f(s - t, t) ds dt = \int_{-\infty}^{z} \int_{-\infty}^{\infty} f(s - t, t) dt ds.$$

Logo, $\int_{-\infty}^{\infty} f(s - t, t) dt$ é a densidade da soma $Z = X + Y$, ou seja,

$$f_Z(z) = \int_{-\infty}^{\infty} f(z - t, t) dt = \int_{-\infty}^{\infty} f(s, z - s) ds,$$

onde foi feita a troca de variáveis $s = z - t$ para obter a última expressão. ■

5.6 Aprendendo um pouco mais

O método do Jacobiano é descrito a seguir para funções mais gerais H.

Suponha que $G_0 \subseteq \mathbb{R}^n, G \subseteq \mathbb{R}^n$ sejam regiões abertas, e que $H : G_0 \to G$ seja uma bijeção entre G_0 e G. Logo, existe a função inversa H^{-1} em G, de modo que $\vec{X} = H^{-1}(\vec{Y})$. Suponha ainda que f é a densidade conjunta de $\vec{X}$ e que $P(\vec{X} \in G_0) = 1$. Se as derivadas parciais de H^{-1} existirem e o Jacobiano J de H^{-1} for diferente de zero para todo $\vec{y} \in G$, utiliza-se o teorema da mudança de variáveis e obter que para $B \subseteq G$, B Boreliano, tem-se

$$P(\vec{Y} \in B) = P(\vec{X} \in H^{-1}(B)) = \int \cdots \int_{H^{-1}(B)} f(x_1, \ldots, x_n) dx_1 \cdots dx_n$$

$$= \int \cdots \int_{B} f(H_1^{-1}(y_1, \ldots, y_n), \ldots, H_n^{-1}(y_1, \ldots, y_n))|J| dy_1 \cdots dy_n.$$

Como $P(\vec{Y} \in G) = P(\vec{X} \in H^{-1}(G)) = P(\vec{X} \in G_0) = 1$, então, para todo Boreliano B no $\mathbb{R}^n$,

$$P(\vec{Y} \in B) = P(\vec{Y} \in B \cap G) = \int \cdots \int_{B \cap G} f(H_1^{-1}(y_1, \ldots, y_n), \ldots, H_n^{-1}(y_1, \ldots, y_n))|J| dy_1 \cdots dy_n.$$

Esta última integral é igual a integral sobre o conjunto B da função que toma o valor $f(H_1^{-1}(y_1, \ldots, y_n), \ldots, H_n^{-1}(y_1, \ldots, y_n))|J|$ para $\vec{y} \in G$, e zero no caso contrário. Portanto, pela definição de densidade,

$$f_{\vec{Y}}(y_1, \ldots, y_n) = \begin{cases} f(H_1^{-1}(y_1, \ldots, y_n), \ldots, H_n^{-1}(y_1, \ldots, y_n))|J|, & \text{se } \vec{y} \in G, \\ 0, & \text{caso contrário.} \end{cases}$$

Observações

(i) Note que J é o Jacobiano da função inversa H^{-1}. Em alguns casos pode ser útil obter J a partir do Jacobiano J' da função H através da relação $J = \frac{1}{J'}|_{\vec{x}=H^{-1}(\vec{y})}$.

(ii) Para obter a distribuição de $\vec{Y} = H(\vec{X})$ quando a dimensão de $\vec{Y}$ é menor que a dimensão de $\vec{X}$ muitas vezes é possível definir outras variáveis aleatórias $Y_1', \ldots, Y_m'$, utilizar o método do Jacobiano para determinar a densidade conjunta de $\vec{Y}, Y_1', \ldots, Y_m'$ e, finalmente, obter a densidade marginal conjunta de $\vec{Y}$. Considere o seguinte exemplo:

Exemplo 5.6.1: Suponha que X_1, X_2 tem densidade conjunta dada por $f(x, y)$ e que o objetivo seja a distribuição de $Y_1 = X_1^2 + X_2$. Como esta não é uma transformação 1-1, ela não possui inversa. Definindo uma nova variável $Y_2 = X_1$ de modo que a função $(Y_1, Y_2) = H(X_1, X_2) = (X_1^2 + X_2, X_1)$ possua uma função inversa diferenciável, $(X_1, X_2) = H^{-1}(Y_1, Y_2) = (Y_2, Y_1 - Y_2^2)$. Deste modo,

$$J = det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -2y_2 \end{pmatrix} = -1$$

Então, $f_{Y_1,Y_2}(y_1, y_2) = f(y_2, y_1 - y_2^2)$. Finalmente, para encontrar f_{Y_1} integra-se sobre todos os possíveis valores da variável Y_2 introduzida:

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f(y_2, y_1 - y_2^2) dy_2.$$

■

(iii) Pode-se utilizar o método do Jacobiano em outros casos em que a função H não é 1-1. Para tanto, suponha que $G, G_1, \ldots, G_k$ sejam subregiões abertas do $\mathbb{R}^n$ tais que $G_1, \ldots, G_k$ sejam disjuntas e $P(\vec{X} \in \cup_{i=1}^{k} G_i) = 1$, tais que a função $H|_{G_l}$, a restrição de H a G_l, seja um correspondência 1-1 entre G_l e G, para $l = 1, \ldots, k$. Suponha que para todo l, a função inversa de $H|_{G_l}$ satisfça as hipóteses do caso anterior, e seja J_l o Jacobiano da inversa de $H|_{G_l}$. Pode-se provar que

$$f_{\vec{Y}}(y_1, \ldots, y_n) = \begin{cases} \sum_{l=1}^{k} f(H|_{G_l}^{-1}(y_1, \ldots, y_n))|J_l|, & \text{se } \vec{y} \in G, \\ 0, & \text{caso contrário.} \end{cases}$$

Para a utilização do método do jacobiano, foi necessário assumir que o vetor $\vec{X}$ possuía densidade conjunta. Na próxima seção será visto como estender este método para um caso mais geral.

5.6.1 Extensão do Método Jacobiano para o Cálculo de Densidades de Funções de Vetores Aleatórios Quaisquer

A extensão supõe apenas que existe pelo menos uma variável no vetor $\vec{X}$ que é absolutamente contínua dado os valores das demais variáveis em $\vec{X}$.

Para um dado vetor $\vec{z} \in \mathbb{R}^m$, sejam G_0 e $G^{\vec{z}}$ regiões abertas do $\mathbb{R}^n$, e $g : G_0 \times \{\vec{z}\} \to G^{\vec{z}} \times \{\vec{z}\}$ uma função bijetiva. Seja $f_{\vec{X}|\vec{Z}}$ a densidade condicional conjunta do vetor aleatório $\vec{X} = (X_1, \ldots, X_n)$ dado o vetor aleatório $\vec{Z} = (Z_1, \ldots, Z_m)$, onde $P((X_1, \ldots, X_n) \in G_0|\vec{Z} = \vec{z}) = 1$. Não assume-se qualquer hipótese sobre o tipo do vetor $\vec{Z}$, o qual pode ter partes discreta, contínua ou singular diferentes de zero.

Sejam $Y_1, \ldots, Y_n$ variáveis obtidas a partir de funções dos vetores $\vec{X}$ e $\vec{Z}$, i.e., $Y_i = g_i(X_1, \ldots, X_n, Z_1, \ldots, Z_m)$, $i = 1, 2, \ldots, n$. Portanto, existe função inversa $h = g^{-1}$ definida em $G^{\vec{z}} \times \{\vec{z}\}$, onde

$$X_1 = h_1(Y_1, \ldots, Y_n, z_1, \ldots, z_m), \ldots, X_n = h_n(Y_1, \ldots, Y_n, z_1, \ldots, z_m),$$

e $h_i(Y_1, \ldots, Y_n, z_1, \ldots, z_m) = z_i$, para $i \in \{n + 1, n + 2, \ldots, n + m\}$.

Suponha que existam as derivadas parciais

$$\frac{\partial X_i}{\partial Y_j} = \frac{\partial h_i(Y_1, \ldots, Y_n, z_1, \ldots, z_m)}{\partial Y_j},$$

para $i, j \in \{1, \ldots, n\}$ e que elas sejam contínuas em $G^{\vec{z}} \times \{\vec{z}\}$. Define-se o jacobiano condicional dado $\vec{Z} = \vec{z}$ como $J(\vec{X}, \vec{Y}|\vec{Z} = \vec{z})$ pelo determinante:

$$J(\vec{X}, \vec{Y}|\vec{Z} = \vec{z}) = det \begin{pmatrix} \frac{\partial X_1}{\partial Y_1} & \cdots & \frac{\partial X_1}{\partial Y_n} \\ \vdots & & \vdots \\ \frac{\partial X_n}{\partial Y_1} & \cdots & \frac{\partial X_n}{\partial Y_n} \end{pmatrix}$$

Suponha que $J(\vec{X}, \vec{Y}|\vec{Z} = \vec{z})$ seja diferente de zero para todo $\vec{Y} \in G^{\vec{z}}$. Então para $B \subseteq G^{\vec{z}}$, B boreliano, seja $h(B \times \{\vec{z}\}) = \{(x_1, \ldots, x_n) :$ para algum $\vec{y} \in B, x_i = h_i(\vec{y}, \vec{z})$ para todo $i = 1, \ldots, n\}$. Utilizando o teorema de mudança de variáveis, tem-se

$$\begin{aligned} & P(\vec{Y} \in B|\vec{Z} = \vec{z}) = P(\vec{X} \in h(B \times \{\vec{z}\})|\vec{Z} = \vec{z}) \\ & = \int \cdots \int_{h(B \times \{\vec{z}\})} f_{\vec{X}|\vec{Z}}(x_1, \ldots, x_n|\vec{z}) dx_1 \cdots dx_n \\ & = \int \cdots \int_{B} f_{\vec{X}|\vec{Z}}(h_1(\vec{y}, \vec{z}), \ldots, h_n(\vec{y}, \vec{z})|\vec{z}) |J(\vec{x}, \vec{y}|\vec{Z} = \vec{z})| dy_1 \cdots dy_n. \end{aligned}$$

Como $P(\vec{Y} \in G^{\vec{z}}|\vec{Z} = \vec{z}) = P(\vec{X} \in h(G^{\vec{z}} \times \{\vec{z}\})|\vec{Z} = \vec{z}) = P(\vec{X} \in G_0|\vec{Z} = \vec{z}) = 1$, tem-se que para todo boreliano B no $\mathbb{R}^n$,

$$\begin{aligned} & P(\vec{Y} \in B|\vec{Z} = \vec{z}) = P(\vec{Y} \in B \cap G^{\vec{z}}|\vec{Z} = \vec{z}) \\ & = \int \cdots \int_{B \cap G^{\vec{z}}} f_{\vec{X}|\vec{Z}}(h_1(\vec{y}, \vec{z}), \ldots, h_n^{(}\vec{y}, \vec{z})|\vec{z}) |J(\vec{x}, \vec{y}|\vec{Z} = \vec{z})| dy_1 \cdots dy_n. \end{aligned}$$

Esta última integral é igual a integral sobre o conjunto B da função que toma o valor $f_{\vec{X}|\vec{Z}}(h_1(\vec{y}, \vec{z}), \ldots, h_n(\vec{y}, \vec{z})|\vec{z}) |J(\vec{x}, \vec{y}|\vec{Z} = \vec{z})|$ para $\vec{y} \in G^{\vec{z}}$, e zero, caso contrário. Portanto, pela definição de densidade condicional:

$$\begin{aligned} & f_{\vec{Y}|\vec{Z}}(y_1, \ldots, y_n|\vec{z}) \\ & = \begin{cases} f_{\vec{X}|\vec{Z}}(h_1(\vec{y}, \vec{z}), \ldots, h_n(\vec{y}, \vec{z})|\vec{z}) |J(\vec{x}, \vec{y}|\vec{Z} = \vec{z})|, & \text{se } \vec{y} \in G^{\vec{z}}, \\ 0, \text{ caso contrário.} \end{cases} \end{aligned}$$

A fim de se obter a densidade incondicional do vetor $\vec{Y}$, calcula-se a esperança[2] da densidade condicional $f_{\vec{Y}|\vec{Z}}$ com respeito a distribuição do vetor aleatório $\vec{Z}$. Portanto,

[2]este conceito será dado no próximo capítulo, mas nesta seção ...

$$f_{\vec{Y}}(\vec{y}) = \int f_{\vec{Y}|\vec{Z}}(y_1, \ldots, y_n|\vec{z})dF_{\vec{Z}}(\vec{z}).$$

No caso particular em que $\vec{Z}$ for um vetor aleatório com densidade conjunta $f_{\vec{Z}}$,

$$f_{\vec{Y}}(\vec{y}) = \int \cdots \int f_{\vec{Y}|\vec{Z}}(y_1, \ldots, y_n|\vec{z})f_{\vec{Z}}(\vec{z})dz_1 \cdots dz_m,$$

e, no caso particular em que $\vec{Z}$ for um vetor aleatório discreto com função probabilidade de massa conjunta $p_{\vec{Z}}$,

$$f_{\vec{Y}}(\vec{y}) = \sum_{\vec{z}} f_{\vec{Y}|\vec{Z}}(y_1, \ldots, y_n|\vec{z})p_{\vec{Z}}(\vec{z}).$$

Exemplo 5.6.2: Suponha que X_1 é uma variável aleatória discreta que assume os valores $10, 15, 20$ com probabilidades $1/4$, $1/2$, e $1/4$, respectivamente. Sejam ainda X_2 e X_3 variáveis aleatórias que são condicionalmente independentes dado X_1 e com distribuições condicionais $X_2|X_1 = k \sim Exp(k)$ e $X_3|X_1 = k \sim Exp(2k)$. Seja $Y = X_1^2 + X_2^2 + X_3^2$ e $Z = arctg(\frac{X_2}{X_3})$. Determinar a densidade conjunta de (Y, Z).
Solução: A densidade condicional conjunta de $(X_2, X_3)|X_1 = k$ é dada por

$$2k^2 e^{-kx_2-2kx_3}U(x_2)U(x_3).$$

Tem-se que $X_1 = k$,

$$P((Y, Z) \in [k^2, \infty) \times [0, \pi/2]) = 1,$$

$$X_2 = (\frac{Y - k^2}{2})^{\frac{1}{2}} \text{sen}\, Z,$$

$$X_3 = (\frac{Y - k^2}{2})^{\frac{1}{2}} \cos Z.$$

Portanto, o Jacobiano condicional dado que $X_1 = k$ é dado por:

$$J((X_2, X_3), (Y, Z)|X_1 = k) = det \begin{pmatrix} \frac{\text{sen}\, Z}{4}(\frac{Y-k^2}{2})^{-\frac{1}{2}} & \cos Z(\frac{Y-k^2}{2})^{\frac{1}{2}} \\ \frac{\cos Z}{4}(\frac{Y-k^2}{2})^{-\frac{1}{2}} & -\text{sen}\, Z(\frac{Y-k^2}{2})^{\frac{1}{2}} \end{pmatrix} = -\frac{1}{4}.$$

Assim, a densidade condicional de (Y, Z) dado que $X_1 = k$ é dada por:

$$f_{Y,Z|X_1}(y, z|k)$$
$$= \begin{cases} f_{X_2,X_3|X_1}(\text{sen}\, z(\frac{y-k^2}{2})^{\frac{1}{2}}, \cos z(\frac{y-k^2}{2})^{\frac{1}{2}}|k)\frac{1}{4}, & (y, z) \in [k^2, \infty) \times [0, \pi/2), \\ 0, & \text{caso contrário.} \end{cases}$$
$$= \begin{cases} \frac{k^2}{2} \exp(-k(\text{sen}\, z + 2\cos z)(\frac{y-k^2}{2})^{\frac{1}{2}}), & (y, z) \in [k^2, \infty) \times [0, \pi/2), \\ 0, & \text{caso contrário.} \end{cases}$$

Calculando a esperança (que será vista em capítulo posterior) em termos da distribuição de X_1, tem-se:

$$\begin{aligned} &f_{Y,Z}(y,z) = P(X_1 = 10) f_{Y,Z|X_1}(y,z|10) \\ &+P(X_1 = 15) f_{Y,Z|X_1}(y,z|15) + P(X_1 = 20) f_{Y,Z|X_1}(y,z|20), \end{aligned}$$

ou seja,

$$f_{Y,Z}(y,z)$$
$$= \begin{cases} \frac{1}{4}(50\exp(-10(\operatorname{sen} z + 2\cos z)(\frac{y-100}{2})^{\frac{1}{2}})), \quad \text{se } (y,z) \in [100,225) \times [0,\pi/2); \\ \frac{1}{4}(50\exp(-10(\operatorname{sen} z + 2\cos z)(\frac{y-100}{2})^{\frac{1}{2}})) + \frac{1}{2}(\frac{225}{2}\exp(-15(\operatorname{sen} z + 2\cos z)(\frac{y-225}{2})^{\frac{1}{2}})), \\ \text{se } (y,z) \in [225,400) \times [0,\pi/2); \\ \frac{1}{4}(50\exp(-10(\operatorname{sen} z + 2\cos z)(\frac{y-100}{2})^{\frac{1}{2}})) + \frac{1}{2}(\frac{225}{2}\exp(-15(\operatorname{sen} z + 2\cos z)(\frac{y-225}{2})^{\frac{1}{2}})) + \\ \frac{1}{4}(200\exp(-20(\operatorname{sen} z + 2\cos z)(\frac{y-400}{2})^{\frac{1}{2}})), \quad \text{se } (y,z) \in [400,\infty) \times [0,\pi/2); \\ 0, \quad \text{caso contrário.} \end{cases}$$

■

Observações:

(i) No desenvolvimento na seção anterior, para obter a distribuição de $\vec{Y} = g(\vec{X}, \vec{Z})$ assumiu-se que o vetor $\vec{Y}$ tem dimensão igual a dimensão do vetor $\vec{X}$. Quando a dimensão de $\vec{Y}$ é menor que a dimensão de $\vec{X}$, o tratamento é análogo ao caso da utilização do método do Jacobiano para vetores absolutamente contínuos, ou seja, muitas vezes é possível definir outras variáveis aleatórias auxiliares $Y'_1, \ldots, Y'_m$, utilizar a extensão do método do Jacobiano para determinar a densidade condicional conjunta de $\vec{Y}, Y'_1, \ldots, Y'_m$ dado $\vec{Z}$ e, finalmente, obter a densidade marginal condicional conjunta de $\vec{Y}$ dado $\vec{Z}$.

(ii) Também pode-se utilizar o método do Jacobiano em outros casos em que a função g não é bijetiva. Para tanto, dado que $\vec{Z} = \vec{z}$, suponha que $G^{\vec{z}}$, $G_1^{\vec{z}}, \ldots, G_k^{\vec{z}}$ sejam subregiões abertas do $\mathbb{R}^n$ tais que $G_1^{\vec{z}}, \ldots, G_k^{\vec{z}}$ sejam disjuntas e $P((\vec{X}, \vec{Z}) \in (\cup_{i=1}^k G_i^{\vec{z}}) \times \{\vec{z}\}) = 1$, tais que a função $g|_{G_l^{\vec{z}}}$, a restrição de g a $G_l^{\vec{z}}$ seja bijetiva entre $G_l^{\vec{z}}$ e $G^{\vec{z}}$, para $l = 1, \ldots, k$. Suponha que para todo l, a função inversa de $g|_{G_l^{\vec{z}}}$ satisfaça as hipóteses do caso anterior, e seja $J_l^{\vec{z}}$ o Jacobiano condicional dado que $\vec{Z} = \vec{z}$ da inversa de $g|_{G_l^{\vec{z}}}$. Pode-se provar que

$$f_{\vec{Y}|\vec{Z}}(y_1, \ldots, y_n|\vec{z}) = \begin{cases} \sum_{l=1}^k f_{\vec{X}|\vec{Z}}(g|_{G_l^{\vec{z}}}^{-1}(y_1, \ldots, y_n, \vec{z})|\vec{z})|J_l^{\vec{z}}|, & \text{se } \vec{y} \in G^{\vec{z}}, \\ 0, \text{caso contrário.} \end{cases}$$

5.7 Exercícios

5.1 Nos itens a seguir, considere que cada variável aleatória é uniformemente distribuída na região dada. Em cada caso, (i) esboçe um gráfico enfatizando a região de interesse, (ii) encontre a conjunta (iii) as marginais e (iv) valide[3] os resultados encontrados.

(a) Área no quarto quadrante limitada por $f_1(x) = -x^2 + 2x - 1$ e $f_2(x) = -2x + 2$;

(b) Área no primeiro quadrante limitada por $f_1(x) = \frac{1}{x}, x \neq 0$ e $f_2(x) = -0.75x^2 + 1.75x$;

(c) Área limitada por $f_1(x) = 1.5x^2 + 0.5x$ e $f_2(x) = 0.5x + 1.5$;

(d) Área entre $(x-1)^2 + y^2 = 2$ e $(x-1)^2 + (y-1)^2 = 1$;

(e) Área no primeiro quadrante limitada por $f_1(x) = \frac{2}{x^2+1}$ e $f_2(x) = \frac{1}{x^2}$, $x \neq 0$;

(f) Maior área limitada por $(x-1)^2 + (y+2)^2 = 1$ e $f_(x) = -x^2 + 2x - 2$;

(g) Área limitada por $f_1(x) = \log_{10}(x+1)$, $f_2(x) = 0.5x - 3.5$ e o eixo dos y;

(h) Área no primeiro quadrante limitada por $f_1(x) = \log_{10}(x)$, $f_2(x) = 0.2x - 1$ e $y = 0$;

(i) Maior área no primeiro quadrante limitada por $x^2 + (y-3)^2 = 9$ e $f_(x) = 3\log_{10}(3x+1)$;

(j) Maior área limitada por $(x+1)^2 + (y-2)^2 = 1$ e $f(x) = x^3 + 2$.

5.2 Suponha que X seja uma variável aleatória contínua com função densidade de probabilidade

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

Para $b > 0$ real, determine:

(a) $F(x \mid 0 < X < b) = P(X \leq x \mid 0 < X < b)$, para todo x real.

(b) $f(x \mid 0 < X < b)$, a função densidade condicional de X, dado que $X \in (0, b)$.

5.3 Suponha que as variáveis aleatórias X e Y têm densidade conjunta

$$f(x,y) = \begin{cases} xye^{-\frac{x^2+y^2}{2}}, & x > 0,\ y > 0, \\ 0, & \text{quaisquer outros valores.} \end{cases}$$

Encontre as densidades marginais.

5.4 Sejam $\Omega = \{\omega_1, \omega_2, \omega_3\}$ e $P(\omega_1) = P(\omega_2) = P(\omega_3) = 1/3$. Definindo X, Y e Z como se segue:

$$X(\omega_1) = 1, X(\omega_2) = 2, X(\omega_3) = 3,$$

$$Y(\omega_1) = 2, Y(\omega_2) = 3, Y(\omega_3) = 1,$$

[3]isto é, mostre que $\int_{-\infty}^{+\infty} f(x,y) = 1$, $\int_{-\infty}^{+\infty} f_X(x) = 1$ e $\int_{-\infty}^{+\infty} f_Y(y) = 1$.

$$Z(\omega_1) = 3, Z(\omega_2) = 1, Z(\omega_3) = 2,$$

mostre que estas três variáveis aleatórias têm a mesma distribuição de probabilidade. Encontre a distribuição de probabilidade de $X + Y$, $Y + Z$ e $X + Z$.

5.5 Suponha que X é uma variável aleatória assumindo os valores $-3, -1, 0, 1, 2, 3, 5, 8$ com as respectivas probabilidades 0.1, 0.2, 0.15, 0.2, 0.1, 0.15, 0.05, 0.05. Determine as probabilidades de:

(a) X ser negativa.

(b) $P(X = -3 \mid X \leq 0)$.

(c) $P(X \geq 3 \mid X > 0)$.

5.6 Considere a variável aleatória bidimensional (X, Y) uniformemente distribuída na região poligonal T de vértices (-2,0), (2,0), (1,1) e (-1,1).

(a) Determine a função de densidade de probabilidade conjunta $f(x, y)$.

(b) Determine a função de densidade de probabilidade marginal $f_X(x)$.

(b) Determine a função de densidade de probabilidade marginal $f_Y(y)$.

(d) Verifique se X e Y são variáveis aleatórias independentes.

5.7 Considere duas variáveis aleatórias X e Y com distribuição de probabilidade conjunta uniforme na região triangular tendo vértices nos pontos (0,0), (0,1) e (1,0).

(a) Escreva a expressão da densidade conjunta.

(b) Determine as densidades marginais.

(c) X e Y são independentes?

5.8 Duas mensagens que estão sendo transmitidas, independentemente uma da outra, podem ser distorcidas ou não. A probabilidade do evento $A = \{$uma mensagem é distorcida$\}$ para a primeira mensagem é p_1 e para a segunda p_2. Seja um sistema de variáveis aleatórias (X, Y) definido como se segue:

$$X = \begin{cases} 1, & \text{se a primeira mensagem é distorcida,} \\ 0, & \text{se a primeira mensagem não é distorcida.} \end{cases}$$

$$Y = \begin{cases} 1, & \text{se a segunda mensagem é distorcida,} \\ 0, & \text{se a segunda mensagem não é distorcida.} \end{cases}$$

(X e Y são os indicadores do evento A).

(a) Encontre a distribuição de probabilidade conjunta do par de variáveis aleatórias (X, Y).

(b) Encontre a função distribuição de probabilidade acumulada $F(x, y)$.

5.9 Sejam duas variáveis aleatórias independentes X e Y, cada uma das quais com distribuição exponencial[4] com diferentes parâmetros. Escreva expressões para

(a) a função densidade conjunta $f(x, y)$ e

(b) a função distribuição conjunta $F(x, y)$.

5.10 Um sistema de variáveis aleatórias (X, Y) tem função densidade conjunta $f(x, y)$. Expresse as seguintes probabilidades em termos de $f(x, y)$:

(a) $\{X > Y\}$;

(b) $\{X >\mid Y \mid\}$;

(c) $\{\mid X \mid > Y\}$;

(d) $\{X - Y > 1\}$.

5.11 Um sistema de variáveis aleatórias (X, Y, Z) tem uma densidade conjunta $f(x, y, z)$. Escreva expressões para:

(a) as densidades $f_X(x)$, $f_Y(y)$

(b) a densidade conjunta $f_{Y,Z}(y, z)$ do vetor aleatório (X, Z);

(c) a densidade condicional $f_{Y,Z}(y, z \mid x)$;

(d) a densidade condicional $f_Y(y \mid x, z)$;

(e) a função de distribuição conjunta $F(x, y, z)$;

(f) a função de distribuição $F_X(x)$ da variável aleatória X;

(g) a função de distribuição $F(x, y)$ do vetor (X, Y).

5.12 Um sistema de variáveis aleatórias (X, Y, Z) se distribui com uma densidade constante no interior de uma bola de raio r. Encontre a probabilidade de que o ponto aleatório (X, Y, Z) caia numa bola concêntrica de raio $r/2$.

5.13 Seja o vetor aleatório (X, Y). Sabe-se que a variável aleatória X segue uma distribuição exponencial com parâmetro λ. Para um dado $X = x > 0$, a variável aleatória Y também segue uma distribuição exponencial com parâmetro x.

(a) Escreva a densidade conjunta $f(x, y)$ de X e Y.

(b) Encontre a densidade de Y.

(c) Encontre a densidade condicional $f_{X|Y}(x \mid y)$.

5.14 Duas pessoas marcam um encontro em um determinado lugar, entre 12:00 e 13:00 horas. Cada uma chega ao local do encontro independentemente e com uma densidade de probabilidade constante no intervalo de tempo assinalado. Encontre a probabilidade de que a primeira pessoa espere não menos que meia hora.

[4] $f_X(x) = \alpha e^{-\alpha x}, x > 0, \alpha > 0$, sendo α o parâmetro.

5.15 Dadas duas variáveis aleatórias X e Y com uma densidade conjunta $f(x,y)$, determine:

(a) a função densidade do máximo das duas variáveis, $Z = \max\{X,Y\}$;

(b) a função densidade do mínimo das duas variáveis, $Z = \min\{X,Y\}$;

(c) a função densidade do máximo e a função densidade do mínimo de n variáveis aleatórias independentes.

5.16 Sejam X e Y variáveis aleatórias discretas e g e h funções tais que satisfaçam a identidade $P(X=x,Y=y) = g(x)h(y)$.

(a) Expresse $P(X=x)$ em termos de g e h.

(b) Expresse $P(Y=y)$ em termos de g e h.

(c) Mostre que $(\sum_x g(x))(\sum_y h(y)) = 1$.

(d) Mostre que X e Y são independentes.

5.17 Sejam X_1 e X_2 duas determinações independentes da variável aleatória X. Encontre a densidade da variável aleatória $Z = X_1/X_2$, onde

$$f_X(x) = \begin{cases} \frac{1000}{x^2}, & x > 1000, \\ 0, & \text{quaisquer outros casos.} \end{cases}$$

5.18 Suponha que as dimensões X e Y de uma chapa retangular de metal possam ser consideradas variáveis aleatórias contínuas independentes com densidades, respectivamente:

$$f_X(x) = \begin{cases} x-1, & 1 < x \le 2, \\ -x+3, & 2 < x < 3, \\ 0, & \text{quaisquer outros casos.} \end{cases}$$

$$f_Y(y) = \begin{cases} 1/2, & 2 < y < 4, \\ 0, & \text{quaisquer outros casos.} \end{cases}$$

Encontre a densidade da área da chapa, $A = XY$.

5.19 Ao mensurar-se T, a duração da vida de uma peça, pode-se cometer um erro, o qual se pode admitir ser uniformemente distribuído sobre (-0.01,0.01). Por isso, o tempo registrado (em horas) pode ser representado por $T + X$, onde T, tem uma distribuição exponencial com parâmetro 0.2 e X tem a distribuição uniforme descrita acima. Determine a densidade de $T + X$, quando T e X forem independentes.

5.20 Sejam T_1 e T_2 variáveis aleatórias independentes com distribuição exponencial de parâmetros λ_1 e λ_2, respectivamente. Encontre a densidade de $M = \max\{T_1,T_2\}$ e de $K = \min\{T_1,T_2\}$.

5.21 As variáveis aleatórias X_i, $i = 1,\ 2$ são mutuamente independentes e seguem uma lei de Poisson[5] com parâmetros λ_i. Mostre que sua soma também segue uma distribuição de Poisson, onde o parâmetro é a soma dos parâmetros.

[5] $P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}, k \in \{0,1,\ldots\}$.

5.22 A intensidade luminosa em um dado ponto é dada por $I = \frac{C}{D^2}$, onde C é o poder luminoso da fonte e D é a distância dessa fonte até o ponto dado. Suponha que C seja uniformemente distribuída sobre (1,2) e D tenha densidade $f_D(d) = e^{-d},\ d > 0$. Encontre a densidade de I admitindo que C e D sejam independentes.

5.23 Dois dados são lançados e os números mostrados nas faces voltadas para cima são anotados como um par ordenado (i, j). Sejam as variáveis aleatórias $X((i, j)) = i + j$ e $Y((i, j)) =\mid i - j \mid$, para $i, j = 1, \ldots, 6$. Determine:

(a) $P(X = 4, Y = 2)$.

(b) Verifique se X e Y são independentes.

Capítulo 6

Esperança e outros Momentos

6.1 Esperança Matemática de uma Variável Aleatória

O conceito de esperança matemática ou valor esperado de uma variável aleatória X, ou a "média" é tão antigo quanto o próprio conceito de probabilidade. Na verdade, é até possível definir probabilidade em termos de esperança, mas esta não é uma maneira comum de se apresentar a teoria, segundo Fine (2012). As seguintes podem ser interpretações da esperança:

(a) Parâmetro m de uma medida de probabilidade, função de distribuição, ou função probabilidade de massa, também conhecido como média.

(b) Operador linear em um conjunto de variáveis aleatórias que retorna um valor típico da variável aleatória interpretado como uma medida de localização da variável aleatória.

(c) Média do resultado de repetidos experimentos independentes no longo prazo.

(d) Preço justo de um jogo com pagamentos descritos por X.

A definição de esperança pode ser motivada considerando o cálculo do resultado médio de 1000 lançamentos de um dado. Uma maneira de calcular este resultado médio seria somar todos os resultados e dividir por 1000. Uma maneira alternativa seria calcular a fração $p(k)$, $k = 1, \ldots, 6$ de todos os lançamentos que tiveram resultado igual a k e calcular o resultado médio através da soma ponderada:

$$1p(1) + 2p(2) + 3p(3) + 4p(4) + 5p(5) + 6p(6).$$

Quando o número de lançamentos torna-se grande as frações de ocorrência dos resultados tendem à probabilidade de cada resultado.

Em geral, define-se a esperança de uma variável discreta como uma soma ponderada onde as probabilidades são os pesos de ponderação.

Definição 6.1.1: Se X é uma variável aleatória *discreta* com valores $\{x_1, x_2, x_3, \ldots\}$ e probabilidades $\{p_1, p_2, p_3, \ldots\}$, respectivamente, então sua esperança é,

$$E(X) = \sum_i x_i p_i,$$

desde que $\sum_i \mid x_i \mid p_i < \infty$. Como $p_i = P(X = x_i)$, então

$$E(X) = \sum_i x_i P(X = x_i).$$

Exemplo 6.1.2: Considere uma variável aleatória X tal que: $P(X = -2) = 0.25$, $P(X = 0) = 0.5$ e $P(X = 2) = 0.25$. Então,

$$E(X) = -2(0.25) + 0(0.5) + 2(0.25) = 0.$$

■

Exemplo 6.1.3: Seja uma variável aleatória X tal que: $P(X = -a) = P(X = a) = 1/3$ e $P(X = b) = 1/3$. Então,

$$E(X) = -a(1/3) + a(1/3) + b(1/3) = b/3.$$

■

Note então que muitas variáveis aleatórias diferentes podem ter o mesmo valor esperado ou esperança. (É só variar o valor de a no exemplo anterior.)

Exemplo 6.1.4: Se $X \in \{1, 2, \ldots, n\}$ for uma variável aleatória com distribuição de probabilidade aleatória com parâmetro n,

$$P(X = k) = \frac{1}{n},\ k = 1, 2, \ldots, n,$$

sua esperança é dada por:

$$E(X) = \sum_{k=1}^{n} kp(k) = \sum_{k}^{n} k\frac{1}{n} = \frac{1}{n}\sum_{k}^{n} k = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}.$$

■

Definição 6.1.5: Se X é uma variável aleatória contínua com densidade $f_X(x)$ então,

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

se $\int_{-\infty}^{+\infty} \mid x \mid f_X(x) dx < \infty$.

Exemplo 6.1.6: Se $f_X(x) = \frac{1}{2}$, $2 < x < 4$, então

$$E(X) = \int_2^4 x\frac{1}{2} dx = 3.$$

■

A esperança de uma variável aleatória mista pode ser calculada pela soma da esperança devido à parcela discreta da distribuição e a esperança com respeito à parcela contínua da distribuição. Para uma definição geral de esperança ver James (1996).

6.2 Propriedades da Esperança

As seguintes propriedades são aplicaçẽs imediatas da definição de esperança:

(i) $P(X = c) = 1 \Rightarrow E(X) = c$.

(ii) $P(X \geq 0) = 1 \Rightarrow E(X) \geq 0$.

(iii) $E(aX) = aE(X)$, onde a um número real qualquer.

Esta propriedade segue facilmente da expressão da esperança de uma função de variável aleatória.

(iv) $E(X + Y) = E(X) + E(Y)$.

No caso discreto,

$$E(X+Y) = \sum_i \sum_j (x_i + y_j)p(x_i, y_j) = \sum_i x_i \sum_j p(x_i, y_j) + \sum_i \sum_j y_j p(x_i, y_j)$$
$$= \sum_i x_i p(x_i) + \sum_j y_j \sum_i p(x_i, y_j) = E(X) + \sum_j y_j p(y_j) = E(X) + E(Y).$$

No caso contínuo,

$$E(X+Y) = E(\varphi(X,Y)) = \int\int (x+y) f_{X,Y}(x,y)dxdy,$$

e pela linearidade da integral,

$$E(X+Y) = \int\int x f_{X,Y}(x,y)dxdy + \int\int y f_{X,Y}(x,y)dxdy = E(X) + E(Y).$$

(v) $E(\sum_i^n a_i X_i) = \sum_i^n a_i E(X_i)$.

Para provar esta propriedade basta usar as duas últimas propriedades e indução matemática.

(vi) $P(X \geq Y) = 1 \Rightarrow E(X) \geq E(Y)$.

Esta segue das Propriedades (ii) e (v), pois

$$P(X \geq Y) = P(X - Y \geq 0),$$

o que, pela Propriedade (ii), implica que $E(X - Y) \geq 0$. Pela Propriedade (v), $E(X - Y) = E(X) - E(Y)$, ou seja pode-se concluir que $E(X) - E(Y) \geq 0$.

(vii) Se $\{X_1, \ldots, X_n\}$ são variáveis aleatórias mutuamente independentes, então

$$E(\prod_{i=1}^n X_i) = \prod_{i=1}^n E(X_i).$$

No caso discreto,

$$\begin{aligned} E(\prod_{i=1}^{n} X_i) &= \sum_{i_1} \ldots \sum_{i_n} x_{i_1} \ldots x_{i_n} p(x_{i_1}, \ldots, x_{i_n}) \\ &= \sum_{i_1} \ldots \sum_{i_n} x_{i_1} \ldots x_{i_n} \prod_{j=1}^{n} p(x_{i_j}) \\ &= \sum_{i_1} x_{i_1} p(x_{i_1}) \ldots \sum_{i_n} x_{i_n} p(x_{i_n}) \\ &= \prod_{i=1}^{n} E(X_i). \end{aligned}$$

No caso contínuo $f_{\vec{X}}(\vec{x}) = \prod_{i=1}^{n} f_{X_i}(x_i)$, logo

$$E(\prod_{i=1}^{n} X_i) = \int \cdots \int x_1 \cdots x_n f_{\vec{X}}(\vec{x}) dx_1 \cdots dx_n$$

$$= \int \cdots \int \prod_{i=1}^{n} x_i f_{X_i}(x_i) dx_1 \cdots dx_n = \prod_{i=1}^{n} \int x_i f_{X_i}(x_i) dx_i = \prod_{i=1}^{n} E(X_i).$$

De maneira análoga, pode-se provar a seguinte generalização deste resultado:
Se $\{X_1, \ldots, X_n\}$ são variáveis aleatórias mutuamente independentes, então

$$E(\prod_{i=1}^{n} G(X_i)) = \prod_{i=1}^{n} E(G(X_i)).$$

(viii) Se Y for uma variável aleatória que assume valores inteiros não-negativos, então

$$E(Y) = \sum_{k=1}^{\infty} k P(Y = k) = \sum_{k=1}^{\infty} \sum_{j=1}^{k} P(Y = k),$$

trocando a ordem dos somatórios:

$$E(Y) = \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} P(Y = k) = \sum_{j=1}^{\infty} P(Y \geq j).$$

6.3 Esperança Condicional

O conceito da esperança condicional de X dado $Y = y$, $E(X \mid Y = y)$, deve ser entendido como uma média da variável X quando um valor de Y está fixado.

Definição 6.3.1: Se (X, Y) for um vetor aleatório discreto, então a esperança condicional de X dado $Y = y_j$ é

$$E(X \mid Y = y_j) = \sum_i x_i P(X = x_i \mid Y = y_j),$$

desde que a série seja absolutamente convergente e onde y_j é um valor fixo no contradomínio de Y.

Exemplo 6.3.2: Considere que a distribuição de probabilidade conjunta do vetor aleatório (X, Y) é dada por

$$\begin{aligned} p(1,1) &= \frac{1}{9}, \quad p(2,1) = \tfrac{1}{3}, \quad p(3,1) = \frac{1}{9}, \\ p(1,2) &= \frac{1}{9}, \quad p(2,2) = 0, \quad p(3,2) = \frac{1}{18}, \\ p(1,3) &= 0, \quad p(2,3) = \tfrac{1}{6}, \quad p(3,3) = \frac{1}{9}. \end{aligned}$$

Calcular $E(X \mid Y = j)$, para $j = 1, 2, 3$.
Solução. Calcula-se a marginal de Y, em seguida as condicionais e por fim a esperança pedida.

$$P(Y = 1) = \frac{5}{9},\ P(Y = 2) = \frac{3}{18},\ P(Y = 3) = \frac{5}{18}.$$

$$\begin{aligned} P(X = 1 \mid Y = 1) &= \frac{1}{5}, \quad P(X = 1 \mid Y = 2) = \tfrac{2}{3}, \quad P(X = 1 \mid Y = 3) = 0, \\ P(X = 2 \mid Y = 1) &= \frac{3}{5}, \quad P(X = 2 \mid Y = 2) = 0, \quad P(X = 2 \mid Y = 3) = \frac{3}{5}, \\ P(X = 3 \mid Y = 1) &= \frac{1}{5}, \quad P(X = 3 \mid Y = 2) = \tfrac{1}{3}, \quad P(X = 3 \mid Y = 3) = \frac{2}{5}. \end{aligned}$$

Logo,

$$E(X \mid Y = 1) = \sum_{i=1}^{3} x_i P(X = x_i \mid Y = 1) = 1 \times \frac{1}{5} + 2 \times \frac{3}{5} + 3 \times \frac{1}{5} = 2.$$

De forma similar são calculadas $E(X \mid Y = 2)$ e $E(X \mid Y = 3)$. ■

Definição 6.3.3: Se (X, Y) for um vetor aleatório contínuo, então a esperança condicional de X dado $Y = y$ é

$$E(X \mid Y = y) \int_{-\infty}^{+\infty} x f(x \mid y) dx,$$

se a integral for absolutamente convergente.

Exemplo 6.3.4: Seja a densidade conjunta do vetor (X, Y) dada por

$$f(x,y) = \begin{cases} 6xy(2-x-y), & 0 < x < 1, 0 < y < 1, \\ 0, & \text{caso contrário}, \end{cases}$$

Calcular o que acontece, em média, com a variável X quando $Y = y$, onde $y \in (0,1)$, isto é, calcular $E(X \mid Y = y)$.
Solução. Inicialmente é preciso calcular $f_Y(y)$, em seguida $f(x \mid y)$ para, por fim, computar a esperança desejada.

$$f_Y(y) = \int_0^1 6xy(2-x-y)dx = y(4-3y).$$

Logo,

$$f_Y(y) = \begin{cases} y(4-3y), & 0 < y < 1, \\ 0, & \text{caso contrário}, \end{cases}$$

Portanto,

$$f(x \mid y) = \frac{6xy(2-x-y)}{y(4-3y)} = \frac{6x(2-x-y)}{4-3y}, 0 < x < 1, 0 < y < 1.$$

Por fim,

$$E(X \mid Y = y) = \int_0^1 x\frac{6x(2-x-y)}{4-3y}dx = \frac{5-4y}{8-6y}, 0 < y < 1.$$

■

6.4 Esperança Matemática de Funções e Vetores de Variáveis Aleatórias

Se X for uma variável aleatória e se $Y = H(X)$, então Y também será uma variável aleatória. Consequentemente, pode-se calcular $E(Y)$. Existem duas maneiras equivalentes de calcular $E(Y)$, quer a variável seja discreta, quer seja contínua: (i) primeiro, encontrar a lei de probabilidade da variável $Y = H(X)$ pelos métodos já vistos anteriormente para, em seguida, calcular a esperança da variável Y; (ii) calcular a esperança de Y diretamente usando a função $H(X)$.

Se $\vec{X}$ for um vetor aleatório e $Y = H(\vec{X})$ os comentários anteriores valem, isto é, pode-se calcular $E(Y)$ a partir do cálculo inicial de $f_Y(y)$ ou calcular $E(Y)$ usando $H(\vec{X})$.

6.4.1 Caso Discreto

Definição 6.4.1: Seja X uma variável aleatória discreta e seja $Y = H(X)$. Se Y assumir os seguintes valores $y_1, y_2, \ldots$ e se $P(Y = y_i) = p(y_i)$, define-se:

$$E(Y) = \sum_{i=1}^{\infty} y_i p(y_i).$$

Exemplo 6.4.2: Suponha X tal que $P(X = k) = \binom{2}{k}(\frac{1}{2})^k(\frac{1}{2})^{2-k}$, para $k = 0, 1, 2$, e $Y = H(X) = X + 1$. Calcular $E(Y)$.
Solução. O contradomínio de Y é $R_Y = \{1, 2, 3\}$ porque $X = 0 \Rightarrow Y = 1$, $X = 1 \Rightarrow Y = 2$ e $X = 2 \Rightarrow Y = 3$. Portanto, $P(Y = 1) = P(X = 0) = \frac{1}{4}$, $P(Y = 2) = P(X = 1) = \frac{2}{4}$ e $P(Y = 3) = P(X = 2) = \frac{1}{4}$. Assim,

$$E(Y) = 1 \times \frac{1}{4} + 2 \times \frac{2}{4} + 3 \times \frac{1}{4} = 2.$$

■

Conforme visto no capítulo anterior pode-se determinar as probabilidades $p(y_i)$ dado que sabe-se a distribuição de X. No entanto, é possível encontrar $E(Y)$ sem, preliminarmente, encontrar a distribuição de probabilidade de Y, partindo-se apenas do conhecimento da distribuição de probabilidade de X, conforme mostra o seguinte teorema.

Teorema 6.4.3: *Seja X uma variável aleatória discreta assumindo os valores $x_1, x_2, \ldots$ e seja $Y = H(X)$. Se $p(x_i) = P(X = x_i)$, então*

$$E(Y) = E(H(X)) = \sum_{i=1}^{\infty} H(x_i)p(x_i).$$

Prova: Reordenando o somatório $\sum_{i=1}^{\infty} H(x_i)p(x_i)$, e agrupando os termos onde x_i tem a mesma imagem de acordo com a função H, ou seja, sejam $x_{i1}, x_{i2}, \ldots$, todos os valores x_i tal que $H(x_{ij}) = y_i$ para $j \geq 1$, onde $y_1, y_2, \ldots$ são os possíveis valores de Y, tem-se,

$$\sum_{i=1}^{\infty} H(x_i)p(x_i) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} H(x_{ij})p(x_{ij}) = \sum_{i=1}^{\infty} y_i \sum_{j=1}^{\infty} p(x_{ij}) = \sum_{i=1}^{\infty} y_i p(y_i) = E(Y).$$

▮

Exemplo 6.4.4: Sejam X e Y como no Exemplo 11.4.3. Calculando $E(Y)$ sem encontrar a distribuição de probabilidade de Y.
Solução.

$$E(Y) = E(X + 1) = \sum_{k=0}^{2}(k + 1)P(X = k) = \sum_{k=0}^{2} kP(X = k) + \sum_{k=0}^{2} P(X = k) = 1 + 1 = 2.$$

Lembrando que $\sum_{k=0}^{2} P(X = k) = 1$.

O cálculo de $E(Y)$ também poderia ter sido realizado usando propriedades da esperança, da seguinte forma:

$$E(Y) = E(X + 1) = E(X) + E(1) = E(X) + 1 = 1 + 1 = 2.$$

■

Definição 6.4.5: Se $\vec{X}$ um vetor aleatório com distribuição de probabilidade $p_{\vec{X}}$ e $Y = H(\vec{X})$ assume os valores $y_1, y_2, \ldots$, de forma que $P(Y = y_i) = p(y_i)$, então,

$$E(Y) = E(H(\vec{X})) = \sum_i y_i p_{\vec{Y}}(y_i).$$

Exemplo 6.4.6: Seja $\vec{X} = (X_1, X_2)$ com $R_{\vec{X}} = \{(1,1),(1,2),(2,1),(2,2)\}$, $P(X_1 = i, X_2 = j) = \frac{1}{4}$, para $i, j = 1, 2$ e $Y = H(\vec{X}) = X_1 + X_2$. Calcular $E(Y) = E(H(\vec{X}))$.
Solução. $R_Y = \{2,3,4\}$ e $P(Y = 2) = P(X_1 = 1, X_2 = 1) = \frac{1}{4}$, $P(Y = 3) = P(X_1 = 1, X_2 = 2) + P(X_1 = 2, X_2 = 1) = \frac{2}{4}$ e $P(Y = 4) = P(X_1 = 2, X_2 = 2) = \frac{1}{4}$. Logo,

$$E(Y) = 2 \times \frac{1}{4} + 3 \times \frac{2}{4} + 4 \times \frac{1}{4} = 3.$$

■

Teorema 6.4.7: *Seja $\vec{X}$ um vetor aleatório com distribuição de probabilidade $p_{\vec{X}}(\vec{x_i}) = P(X_1 = x_{1i}, X_2 = x_{2j}, \ldots, X_n = x_{nk})$ e $Y = H(\vec{X})$. Então,*

$$E(Y) = E(H(\vec{X})) = \sum_n \ldots \sum_j \sum_i H(\vec{x_i}) p_{\vec{X}}(\vec{x_i}),$$

em que os $\vec{x_i}$ são os valores assumidos pelo vetor aleatório $\vec{X}$.

Exemplo 6.4.8: Suponha $\vec{X}$ como no Exemplo 6.4.6.
Solução. Então, o cálculo de $E(Y)$ é dado por

$$\begin{aligned}
E(Y) &= E(X_1 + X_2) = \sum_j \sum_i (i + j) P(X_1 = i, X_2 = j) \\
&= \sum_j \sum_i (iP(X_1 = i, X_2 = j) + jP(X_1 = i, X_2 = j)) \\
&= \sum_j \sum_i (iP(X_1 = i, X_2 = j)) + \sum_j \sum_i (jP(X_1 = i, X_2 = j) \\
&= \sum_j (\sum_i (iP(X_1 = i, X_2 = j))) + \sum_j (\sum_i (jP(X_1 = i, X_2 = j)) \\
&= \sum_j (P(X_1 = 1, X_2 = j) + 2P(X_1 = 2, X_2 = j)) \\
&\quad + \sum_j (jP(X_1 = 1, X_2 = j) + jP(X_1 = 2, X_2 = j)) \\
&= P(X_1 = 1, X_2 = 1) + P(X_1 = 1, X_2 = 2) + 2(P(X_1 = 2, X_2 = 1) \\
&\quad + P(X_1 = 2, X_2 = 2)) + P(X_1 = 1, X_2 = 1) + P(X_1 = 2, X_2 = 1) + \\
&\quad 2(P(X_1 = 2, X_2 = 2) + P(X_1 = 2, X_2 = 2)) = 3.
\end{aligned}$$

■

6.4.2 Caso Contínuo

Definição 6.4.9: Seja X uma variável aleatória contínua e $Y = H(X)$. Então,

$$E(Y) = \int_{-\infty}^{+\infty} y f_Y(y) dy,$$

se $\int_{-\infty}^{+\infty} \mid y \mid f_Y(y) dy < \infty$.

Exemplo 6.4.10: Supondo X com densidade

$$f_X(x) = \begin{cases} 2(1-x), & 0 < x < 1, \\ 0, & \text{caso contrário,} \end{cases}$$

e $Y = H(X) = 2X$. Calcular $E(Y)$.
Solução. Inicialmente será calculada a densidade de Y, e, em seguida, a esperança de Y.

$$F_Y(y) = P(Y \leq y) = P(X \leq \frac{y}{2}) = F_X(\frac{y}{2}) \Rightarrow$$

$$f_Y(y) = f_X(\frac{y}{2}) \times \frac{1}{2}) = 1 - \frac{y}{2}.$$

Logo,

$$f_X(x) = \begin{cases} 1 - \frac{y}{2}, & 0 < y < 2, \\ 0, & \text{caso contrário.} \end{cases}$$

Portanto,

$$E(Y) = \int_0^2 y(1 - \frac{y}{2}) dy = \frac{2}{3}.$$

■

Definição 6.4.11: Sendo $\vec{X}$ um vetor aleatório e $Y = H(\vec{X})$, então

$$E(Y) = \int_{-\infty}^{+\infty} y f_Y(y) dy,$$

onde $f_Y(y)$ foi calculada inicialmente.

Exemplo 6.4.12: Sejam

$$f_X(x) = \begin{cases} 2x, & 0 \leq x \leq 1, \\ 0, & \text{caso contrário,} \end{cases}$$

e

$$f_Y(y) = \begin{cases} \frac{y^2}{9}, & 0 \leq y \leq 3, \\ 0, & \text{caso contrário,} \end{cases}$$

variáveis aleatórias independentes. O objetivo é calcular $E(Z)$ onde $Z = H(X,Y) = XY$. Similar ao Exemplo 6.4.10 primeiro calcula-se a densidade de Z para em seguida calcular $E(Z)$. A densidade de Z vai ser calculada usando o método do Jacobiano.

$$\begin{cases} z = xy & \Rightarrow \quad y = \frac{z}{x} = \frac{z}{u}, \\ u = x & \Rightarrow \quad x = u. \end{cases}$$

Logo o Jacobiano é:

$$J = \begin{vmatrix} \frac{\partial x}{\partial z} \frac{\partial x}{\partial u} \\ \frac{\partial y}{\partial z} \frac{\partial y}{\partial u} \end{vmatrix}$$

Logo $J = \frac{1}{u}$ e então,

$$f(z,u) = f_X(u) f_Y(\frac{z}{u}) |\frac{1}{u}| = \frac{2z^2}{9u^2}.$$

Como $0 \leq x \leq 1$ e $0 \leq y \leq 3$ então $\frac{z}{3} \leq u \leq 1$ e portanto,

$$f_Z(z) = \int_{\frac{z}{3}}^{1} \frac{2z^2}{9u^2} du = \frac{2}{9} z(3-z), \ 0 \leq z \leq 3.$$

Assim,

$$E(Z) = \int_0^3 z \frac{2}{9} z(3-z) dz = \frac{3}{2}.$$

■

As provas dos teoremas a seguir são omitidas desde que fogem ao escopo do livro.

Teorema 6.4.13: *Seja X uma variável aleatória contínua, $Y = H(X)$, então*

$$E(Y) = \int y f_Y(y) dy = \int H(x) f_X(x) dx,$$

desde que estas integrais existam.

Exemplo 6.4.14: Considerando a variável aleatória do Exemplo 6.4.10, a esperança de $Y = 2X$ será calculada sem usar a f_Y.
Solução.

$$E(Y) = \int_0^1 2x2(1-x) dx = \frac{2}{3}.$$

$E(Y)$ também poderia ser calculada usando propriedades da esperança, mas, neste caso seria imprescindível o cálculo de $E(X)$:

$$E(Y) = E(2X) = 2E(X).$$

■

Teorema 6.4.15: *Seja $\vec{X}$ um vetor aleatório e $Y = H(\vec{X})$ uma variável aleatória. Então,*

$$E(Y) = \int y f_Y(y) dy = \int H(\vec{X}) f_{\vec{X}}(\vec{x}) d\vec{x}.$$

Exemplo 6.4.16: Considerando as variáveis do Exemplo 6.4.12:
Solução. $f(x,y) = \frac{2}{9}xy^2,\ 0 \leq x \leq 1,\ 0 \leq y \leq 3$. Logo,

$$E(Z) = \int_0^1 \int_0^3 H(x,y)f(x,y)dxdy = \int_0^1 \int_0^3 xy\frac{2}{9}xy^2 dxdy = \frac{3}{2}.$$

■

Exemplo 6.4.17: A densidade conjunta do vetor (X,Y) é dada por

$$f(x,y) = \begin{cases} \frac{1}{2}ye^{-xy}, & 0 < x < +\infty, 0 < y < 2, \\ 0, & \text{caso contrário,} \end{cases}$$

Calcular $E(e^{X/2} \mid Y = 1)$.
Solução.

$$f_Y(y) = \int_0^\infty \frac{1}{2}ye^{-xy}dx = \frac{1}{2}, 0 < y < 2.$$

$$f(x \mid y) = \frac{f(x,y)}{f_Y(y)} = ye^{-xy}, 0 < x < +\infty, 0 < y < 2.$$

$$E(e^{X/2} \mid Y = 1) = \int_0^\infty e^{\frac{x}{2}}e^{-x}dx = 2.$$

■

6.5 Momentos

Informações parciais sobre a lei de probabilidade de uma variável aleatória X podem ser obtidas através de esperanças matemáticas de potências de X, as quais são chamadas de *momentos de* X.

Definição 6.5.1: Para qualquer inteiro não-negativo n, o n-ésimo momento da variável aleatória X é

$$E(X^n),$$

se esta esperança existe.

Este momento é usualmente denominado de momento em torno do zero, uma vez que poderia ser escrito como $E((X-0)^n)$. Observe que quando $n = 1$ tem-se a esperança matemática de X.

Exemplo 6.5.2: Seja X tal que

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

Então, o segundo momento de X, $E(X^2)$ é:

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k(k-1) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} + \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= n(n-1)p^2 \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} (1-p)^{n-k} + np \\
&= n(n-1)p^2 \sum_{j=0}^{m} \frac{(m)!}{(j)!(m-j)!} p^j (1-p)^{m-j} + np = n(n-1)p^2 + np.
\end{aligned}$$

■

Exemplo 6.5.3: Seja

$$f_X(x) = \begin{cases} 2x, & 0 < x < 1, \\ 0, & \text{caso contrário,} \end{cases}$$

$$E(X) = \int_0^1 x2xdx = \frac{2}{3}.$$

$$E(X^2) = \int_0^1 x^2 2xdx = \frac{1}{2}.$$

■

Teorema 6.5.4: *Se o k-ésimo momento de uma variável aleatória existir, então todos os momentos de ordem menores do que k também existem.*

Prova: Por hipótese, $E(|X^k|) < \infty$, logo $E(1 + |X^k|) < \infty$. Como para qualquer j tal que $0 < j < k$, $|X^j| \leq 1 + |X^k|$, e $1 + |X^k|$ é integrável, tem-se que $|X^j|$ também é integrável, isto é $E(|X^j|) < \infty$. ■

6.5.1 Momentos Centrais. Variância

Definição 6.5.5: Se X é uma variável aleatória seu n-ésimo momento central em torno de $E(X)$ é

$$E(X - E(X))^n,$$

se esta esperança existir.

O primeiro momento central em torno da média é zero, pois

$$E(X - E(X)) = E(X) - E(E(X)) = E(X) - E(X) = 0.$$

Definição 6.5.6: O segundo momento central é a *variância*, $V(X)$, e é dado por

$$V(X) = E(X - E(X))^2.$$

A variância também pode ser calculada por:

$$\begin{aligned}
V(X) &= E(X - E(X))^2 \\
&= E(X^2 - 2XE(X) + (E(X))^2) \\
&= E(X^2) - 2E(XE(X)) + E((E(X))^2) \\
&= E(X^2) - 2(E(X))^2 + (E(X))^2 \\
&= E(X^2) - (E(X))^2 \\
&= E(X^2) - E(X)^2.
\end{aligned}$$

Exemplo 6.5.7: Considerando a variável do Exemplo 6.5.3, tem-se que

$$V(X) = E(X^2) - E(X)^2 = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{6}.$$

■

Do Teorema Binomial tem-se:

$$(X - E(X))^n = \sum_{k=0}^{n} \binom{n}{k} X^k (-E(X))^{n-k}.$$

Logo, da linearidade da esperança,

$$E(X - E(X))^n = \sum_{k=0}^{n} \binom{n}{k} E(X^k)(-E(X))^{n-k}$$

porque $E((-E(X))^{n-k}) = (-E(X))^{n-k}$, pois tem-se a esperança de uma constante e

$$E(X^n) = E(X - E(X) + E(X))^n = \sum_{k=0}^{n} \binom{n}{k} (E(X))^{n-k} E(X - E(X))^k.$$

Portanto, o n-ésimo momento existe se e somente se o n-ésimo momento central existe.

Exemplo 6.5.8: Considere uma variável aleatória X tal que

$$P(X = m - a) = P(X = m + a) = \frac{1}{2} \Rightarrow E(X^k) = \frac{1}{2}[(m - a)^k + (m + a)^k].$$

$$\begin{gathered}
E(X) = m, \\
E(X^2) = \frac{1}{2}(2m^2 + 2a^2) = m^2 + a^2, \\
V(X) = a^2.
\end{gathered}$$

■

Definição 6.5.9: O *desvio-padrão* σ de uma variável aleatória X é definido como a raiz quadrada positiva da variância,

$$\sigma(X) = \sqrt{V(X)}.$$

A variância mede a distância entre os valores que a variável aleatória assume e sua esperança matemática. Entretanto, quando aplicada em problemas práticos surge uma dificuldade: a unidade de medida da esperança matemática difere da unidade de medida da variância. Por exemplo, se os dados estão em milissegundos, ms, a unidade de medida da média é ms, mas a da variância é ms^2. Por esta razão é comum adotar-se o desvio-padrão.

Exemplo 6.5.10: Suponha que a variável aleatória X assume os valores 0 e 10 com iguais probabilidade. Portanto,

$$E(X) = 5,\ E(X^2) = 50,\ V(X) = 25,\ \sigma = 5.$$

Suponha agora uma outra variável aleatória, Y, assumindo os valores 5 e 5 também com iguais probabilidade. Logo,

$$E(Y) = 5,\ E(X^2) = 25,\ V(X) = 0,\ \sigma = 0.$$

Este exemplo trivial enfatiza a importância de se usar o desvio-padrão. Observe que as médias são as mesmas mesmo para valores tão distintos que as variáveis assumem. ■

6.5.2 Propriedades da Variância

(i) $V(X) \geq 0$.

Prova: Pela definição de variância. ▮

(ii) Se $X = c$, $V(X) = 0$.

Prova: $E(X) = c$, logo $V(X) = E(X-c)^2 = E(0) = 0$. ▮

(iii) $V(X+a) = V(X)$, onde a é uma constante real.

Prova:

$$\begin{aligned} V(X+a) &= E(X+a)^2 - (E(X+a))^2 \\ &= E(X^2) + 2aE(X) + a^2 - (E(X))^2 - 2aE(X) - a^2 \\ &= E(X^2) - (E(X))^2 = V(X). \end{aligned}$$

▮

(iv) $V(aX) = a^2V(X)$

Prova:

$$V(aX) = E(aX)^2 - (E(aX))^2 = a^2E(X)^2 - a^2(EX)^2 = a^2V(X).$$

▮

(v) Se X e Y forem variáveis aleatórias mutuamente independentes, então

$$V(X+Y) = V(X) + V(Y).$$

Prova:

$$\begin{aligned} V(X+Y) &= E(X+Y)^2 - (E(X+Y))^2 \\ &= E(X^2 + 2XY + Y^2) - (E(X))^2 - 2E(X)E(Y) - (EY)^2 \\ &= E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 + 2(E(XY) - E(X)E(Y)) \\ &= E(X^2) + E(Y^2) - (E(X))^2 - (E(Y))^2 + 2E(XY) - 2E(X)E(Y) \\ &= V(X) + V(Y). \end{aligned}$$

porque $E(XY) = E(X)E(Y)$. ∎

(vi) Se $X_1, \ldots, X_n$ são variáveis aleatórias independentes, então

$$V(X_1 + \ldots X_n) = V(X_1) + \ldots + V(X_n).$$

Esta propriedade segue da propriedade anterior e da aplicação de indução matemática.

6.6 A Desigualdade de Tchebychev

Corolário 6.6.1: Desigualdade (Original) de Tchebychev. *Seja X uma variável aleatória, então*

$$P(|X - E(X)| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2}.$$

Prova: Seja $A = \{x : |x| \geq \epsilon\}$ e $g(x) = \frac{x^2}{\epsilon^2}$. Note que $g(x) \geq I_A(x)$, então, pelo Corolário 6.8.4, $P(X \in A) = P(|X| \geq \epsilon) \leq \frac{E(X^2)}{\epsilon^2}$. Substituindo X por $X - E(X)$, tem-se $P(|X - E(X)| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2}$. ∎

Esta desigualdade declara que a probabilidade da variável aleatória diferir da sua média por mais do que uma constante qualquer (ε) é menor ou igual do que $\frac{\sigma^2}{\varepsilon^2}$. Portanto, quanto menor a variância, mais agrupados em torno da média estão os dados e, consequentemente, maior a probabilidade de se obter um valor (dos dados) próximo à média.

A desigualdade de Tchebychev é geral no sentido de que não há qualquer hipótese sobre a lei de probabilidade de X. A única restrição é que $\sigma^2 < \infty$.

Exemplo 6.6.2: Sabe-se que o número de itens produzidos por uma fábrica durante uma semana é uma variável aleaória com média 500 e variância 100. O que pode ser dito sobre a probabilidade de que a produção semanal esteja entre 400 e 600?
Solução. Pela desegualdade de Tchebychev,

$$P(|\ X - 500\ | < 100) \geq 1 - \frac{1}{100} = \frac{99}{100}.$$

Portanto, a probabilidade pedida é de pelo menos 0.99. ■

6.7 Momentos Conjuntos

A noção de momentos conjuntos é definida no contexto de vetores aleatórios.

Definição 6.7.1: Seja $\vec{X} = (X_1, X_2, \ldots, X_k)$ um vetor aleatório k-dimensional. Então, os *momentos conjuntos* de $\vec{X}$ são da forma $E(\prod_{i=1}^{k} X_i^{j_i})$, onde j_i's são inteiros positivos, se esta esperança existir.

De forma análoga ao caso unidimensional pode-se definir também *momentos conjuntos centrais*.

No caso bidimensional a *correlação* e a *covariância* são momentos conjuntos; estes medem o grau de dependência linear entre duas variáveis.

Definição 6.7.2: A *covariância* entre duas variáveis aleatórias X e Y é dada por

$$cov(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y).$$

Note que $cov(X, X) = V(X)$. Na prova da Propriedade (v) da variância aparece a expressão $E(XY) - E(X)E(Y)$, o que implica que em geral tem-se que,

$$V(X + Y) = V(X) + V(Y) + 2cov(X, Y).$$

A seguir será vista uma expressão para a variância da soma de n variáveis aleatórias.

Teorema 6.7.3: *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias tais que $V(X_i) < \infty$, então*

$$V(X_1 + \ldots + X_n) = \sum_{i=1}^{n} V(X_i) + 2 \sum_{i<j} cov(X_i, X_j).$$

Prova:

$$\begin{aligned}
V(X_1 + \cdots + X_n) &= E(X_1 + \cdots + X_n - E(X_1 + \cdots + X_n))^2 \\
&= E(\sum_{i=1}^{n} (X_i - E(X_i))^2 \\
&= E(\sum_{i=1}^{n} (X_i - E(X_i))^2 + 2\sum_{i<j} (X_i - E(X_i))(X_j - E(X_j))) \\
&= \sum_{i=1}^{n} V(X_i) + 2\sum_{i<j} cov(X_i, X_j).
\end{aligned}$$

∎

Corolário 6.7.4: *Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias tais que $V(X_i) < \infty$ e $Cov(X_i, X_j) = 0$ para $i \neq j$, então*

$$V(X_1 + \ldots + X_n) = \sum_{i=1}^{n} V(X_i).$$

O próximo teorema trata de importante desigualdade em teoria da probabilidade:

Teorema 6.7.5: Desigualdade de Cauchy-Schwarz

$$(E(XY))^2 \leq E(X^2)E(Y^2).$$

Prova: $(aX+Y)^2 \geq 0 \Rightarrow E(aX+Y)^2 \geq 0 \Rightarrow a^2E(X^2)+2aE(XY)+E(Y^2) \geq 0$. Observa-se que esta equação do segundo grau em a não pode ter duas raízes reais diferentes, pois caso contrário essa expressão seria negativa para os valores entre as raízes. Então, utilizando a regra do discriminante,

$$4(EXY)^2 - 4EX^2EY^2 \leq 0,$$

o teorema está provado. ∎

Corolário 6.7.6: $(cov(X,Y))^2 \leq V(X)V(Y)$.

Prova: Segue do teorema anterior trocando X por $X - E(X)$ e Y por $Y - E(Y)$. ∎

Definição 6.7.7: O *coeficiente de correlação* entre duas variáveis aleatórias X e Y é dado por

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)V(Y)}}.$$

Definição 6.7.8: Duas variáveis são *não-correlacionadas* se $cov(X,Y) = 0$.

Como já foi provado que se X e Y são independentes, então $E(XY) = E(X)E(Y)$, se X e Y são independentes, elas necessariamente são não-correlacionadas. O contrário nem sempre é verdadeiro como o próximo exemplo ilustra.

Exemplo 6.7.9: Se X é uma variável aleatória tal que $P(X = -a) = P(X = a) = 1/2$ e $Y = X^2$, então $E(XY) = -a^3(1/2) + a^3(1/2) = 0$ e $E(X) = -a(1/2) + a(1/2) = 0$. Logo, $E(XY) = E(X)E(Y) = 0$, ou seja, $Cov(X,Y) = 0$. Porém, X e Y não são independentes, pois Y é uma função de X. ∎

Corolário 6.7.10: $|\rho(X,Y)| \leq 1$.

Prova: Imediata a partir do (6.7.6). ∎

O próximo teorema mostra que o módulo do coeficiente de correlação entre duas variáveis é igual a 1 se, e somente se, as variáveis são linearmente dependentes.

Teorema 6.7.11: *Sejam X e Y variáveis aleatórias com variâncias finitas e positivas. Então,*

(i) $\rho(X,Y) = 1$ *se, e somente se,* $P(Y = aX + b) = 1$ *para algum* $a > 0$ *e* $b \in I\!R$.

(ii) $\rho(X,Y) = -1$ *se, e somente se,* $P(Y = aX + b) = 1$ *para algum* $a < 0$ *e* $b \in I\!R$.

Prova:

(i) Como $(\frac{X-E(X)}{\sqrt{V(X)}} - \frac{Y-E(Y)}{\sqrt{V(Y)}})^2 \geq 0$, então,

$$\begin{aligned} 0 &\leq E(\frac{X-E(X)}{\sqrt{V(X)}} - \frac{Y-E(Y)}{\sqrt{V(Y)}})^2 \\ &= E(\frac{X-E(X)}{\sqrt{V(X)}})^2 + E(\frac{Y-E(Y)}{\sqrt{V(Y)}})^2 - \frac{2}{\sqrt{V(X)V(Y)}}E((X-E(X))(Y-E(Y))) \\ &= \frac{V(X)}{V(X)} + \frac{V(Y)}{V(Y)} - \frac{2Cov(X,Y)}{\sqrt{V(X)V(Y)}} = 2 - 2\rho(X,Y). \end{aligned}$$

Se $\rho(X,Y) = 1$, então

$$E(\frac{X-E(X)}{\sqrt{V(X)}} - \frac{Y-E(Y)}{\sqrt{V(Y)}})^2 = 0,$$

o que por sua vez implica que

$$P(\frac{X-E(X)}{\sqrt{V(X)}} = \frac{Y-E(Y)}{\sqrt{V(Y)}}) = 1,$$

em outras palavras,

$$P(Y = E(Y) + \frac{\sqrt{V(Y)}}{\sqrt{V(X)}}(X - E(X))) = 1.$$

(ii) Análoga, substituindo o sinal "+" por "-" na expressão acima.

∎

6.8 Aprendendo um pouco mais

Corolário 6.8.1: *Se X e Y são variáveis aleatórias em $(\Omega, \mathcal{A}, P)$ tais que $E(|X^t|) < \infty$ e $E(|Y^t|) < \infty$, então $E(|X+Y|^t) < \infty$.*

Prova: $|X+Y| \leq |X| + |Y| \leq 2\max(|X|,|Y|)$. Portanto, $|X+Y|^t \leq 2^t \max(|X|^t, |Y|^t) \leq 2^t(|X|^t + |Y|^t)$. Logo, $E(|X+Y|^t) \leq 2^t(E(|X|^t) + E(|Y|^t) < \infty$. ∎

Como $E(|X|^t) < \infty$ então, $E(|aX|^t) < \infty$, $\forall a \in \mathbb{R}$, este corolário mostra que a classe de variáveis aleatórias em $(\Omega, \mathcal{A}, P)$ possuidoras do t-ésimo momento finito é um espaço vetorial ou espaço linear.

Corolário 6.8.2: $V(X) = E(X-\mu)^2 = \min_{c \in \mathbb{R}} E(X-c)^2$.

Prova:

$$(X-c)^2 = (X-\mu+\mu-c)^2 = (X-\mu)^2 + 2(\mu-c)(X-\mu) + (\mu-c)^2,$$

logo

$$E(X-c)^2 = E(X-\mu)^2 + 2(\mu-c)(E(X)-\mu) + (\mu-c)^2 = V(X) + (\mu-c)^2.$$

Portanto, $E(X-c)^2 \geq E(X-\mu)^2, \forall c \in \mathbb{R}$. ∎

Corolário 6.8.3: Desigualdade de Tchebychev Generalizada. *Dado um conjunto A e uma função $g(x)$ tal que $\forall x$, $g(x) \geq I_A(x)$, tem-se que $P(X \in A) \leq \min(1, E(g(X)))$.*

Prova: Pela monotonicidade da esperança, $E(g(X)) \geq E(I_A(X)) = P(X \in A)$. Mas, como a cota superior pode exceder 1, tem-se que $\min(1, E(g(X))) \geq P(X \in A)$. ∎

Corolário 6.8.4: Desigualdade de Markov. *Seja X uma variável aleatória, então para todo $\epsilon > 0$,*

$$P(|X| \geq \epsilon) \leq \frac{E|X|}{\epsilon}.$$

Prova: Escolha $A = \{x : |x| \geq \epsilon\}$ e $g(x) = \frac{|x|}{\epsilon}$. Note que $g(x) \geq I_A(x)$, então $P(|X| \geq \epsilon) \leq \frac{E(|X|)}{\epsilon}$. ∎

Corolário 6.8.5: *Se $Z \geq 0$ e $E(Z) = 0$, então $P(Z = 0) = 1$.*

Prova: $P(Z \geq \frac{1}{n}) \leq nE(Z) = 0$. Como $[Z > 0] = \cup_n [Z \geq \frac{1}{n}]$,

$$P(Z > 0) = P(\cup_n [Z \geq \frac{1}{n}]) \leq \sum_n P(Z \geq \frac{1}{n}) = 0.$$

Portanto, $P(Z = 0) = 1 - P(Z > 0) = 1$. ∎

Este último corolário implica que, quando $V(X) = 0$, ou seja $E(X - E(X))^2 = 0$, então, $P(X = E(X)) = 1$, isto é, X é constante com probabilidade 1.

O próximo teorema apresenta uma nova relação entre momentos conjuntos de variáveis aleatórias, sendo conhecido como *Desigualdade de Hölder*.

Teorema 6.8.6: *Suponha que p e q satisfazem: $p > 1$, $q > 1$, e $\frac{1}{p} + \frac{1}{q} = 1$. Então, se $E(|X|^p) < \infty$ e $E(|X|^q) < \infty$, tem-se que*

$$E(|XY|) \leq (E|X|^p)^{1/p} (E|Y|^q)^{1/q}.$$

Prova: A prova da desigualdade de Hölder utiliza um argumento de convexidade. Como $|X|^p \geq 0$ (resp., $|X|^q \geq 0$), já foi visto que se $E(|X|^p) = 0$, então $P(X = 0) = 1$. Portanto, em ambos os casos $E(|XY|) = 0$ e a desigualdade de Hölder é válida. Considere então o caso em que o lado direito da desigualdade de Hölder é estritamente positivo.

Para $a > 0$ e $b > 0$, existe $s, t \in \mathbb{R}$ tal que

$$a = \exp(\frac{s}{p}) \text{ e } b = \exp(\frac{t}{q}).$$

Como a função exponencial é convexa e $p^{-1} + q^{-1} = 1$, por convexidade,

$$\exp(\frac{s}{p} + \frac{t}{q}) \leq p^{-1}\exp(s) + q^{-1}\exp(t),$$

ou pela definição de s, t

$$ab \leq p^{-1}a^p + q^{-1}b^q.$$

Agora substituindo a por $\frac{|X|}{(E(|X|^p))^{1/p}}$ e b por $\frac{|Y|}{(E(|Y|^q))^{1/q}}$, tem-se

$$\frac{|XY|}{(E(|X|^p))^{1/p}(E(|Y|^q))^{1/q}} \leq p^{-1}(\frac{|X|}{(E(|X|^p))^{1/p}})^p + q^{-1}(\frac{|Y|}{(E(|Y|^q))^{1/q}})^q.$$

Finalmente, tomando o valor esperado,

$$\begin{aligned} \frac{E(|XY|)}{(E(|X|^p))^{1/p}(E(|Y|^q))^{1/q}} &\leq p^{-1}(\frac{E(|X|^p)}{(E((|X|^p)))})^p + q^{-1}(\frac{E|Y|^q}{(E(|Y|^q))})^q \\ &= p^{-1} + q^{-1} = 1. \end{aligned}$$

∎

6.8.1 Interpretação Geométrica da Esperança

Por definição, $E(X) = \int xdF(x)$, ou seja, $E(X)$ é a integral da diferencial xdF. Mas xdF é uma diferencial de área. Para $x > 0$, xdF é uma diferencial da área da região compreendida entre as curvas $x = 0$, $y = 1$, e $y = F(x)$ no plano Euclideano, cuja área total é dada por $\int_0^\infty (1 - F(x))dx$. Para $x < 0$, $-xdF$ é uma diferencial da área da região compreendida entre as curvas $x = 0$, $y = 0$, e $y = F(x)$ no plano Euclideano, cuja área total é dada por $\int_{-\infty}^0 F(x)dx$. Logo, $E(X) = \int_0^\infty (1 - F(x))dx - \int_{-\infty}^0 F(x)dx$.

Formalmente, prova-se isso da seguinte maneira. A prova é dividida em duas etapas: (a) $\int_0^\infty xdF(x) = \int_0^\infty (1-F(x))dx$ e (b) $\int_{-\infty}^0 xdF(x) = -\int_{-\infty}^0 F(x)dx$. Provando (b). Utilizando integração por partes, tem-se que $\forall a < 0$,

$$\int_a^0 xdF(x) = -aF(a) - \int_a^0 F(x)dx = \int_a^0 (F(a) - F(x))dx.$$

Como $F(a) \geq 0$ e $a < 0$,

$$\int_a^0 xdF(x) \geq -\int_a^0 F(x)dx.$$

Como a desigualdade é válida para todo $a < 0$, tomando o limite quando $a \to -\infty$

$$\int_{-\infty}^{0} xdF(x) \geq -\int_{-\infty}^{0} F(x)dx.$$

Por outro lado, seja $\lambda < 0$. Se $a < \lambda$, então

$$\int_{a}^{0} (F(a) - F(x))dx \leq \int_{\lambda}^{0} (F(a) - F(x))dx = F(a)(-\lambda) - \int_{\lambda}^{0} F(x)dx,$$

e portanto, tomando o limite quando $a \to -\infty$,

$$\int_{-\infty}^{0} xdF(x) \leq -\int_{\lambda}^{0} F(x)dx.$$

Como isto é válido para todo $\lambda < 0$, tomando o limite quando $\lambda \to -\infty$,

$$\int_{-\infty}^{0} xdF(x) \leq -\int_{-\infty}^{0} F(x)dx.$$

Para a parte (a), utilizando integração por partes, tem-se que $\forall b > 0$,

$$\int_{0}^{b} xdF(x) = bF(b) - \int_{0}^{b} F(x)dx = \int_{0}^{b} (F(b) - F(x))dx.$$

Como $F(b) \leq 1$ e $1 - F(x) \geq 0$,

$$\int_{0}^{b} xdF(x) = \int_{0}^{b} (F(b) - F(x))dx \leq \int_{0}^{\infty} (1 - F(x))dx.$$

Como a desigualdade é válida para todo $b > 0$, e tomando o limite quando $b \to \infty$

$$\int_{0}^{\infty} xdF(x) \leq \int_{0}^{\infty} (1 - F(x))dx.$$

Por outro lado, seja $\lambda > 0$. Se $b > \lambda$, então

$$\begin{aligned}
\int_{0}^{b} (F(b) - F(x))dx &\geq \int_{0}^{\lambda} (F(b) - F(x))dx \\
&= \int_{0}^{\lambda} (F(b) - 1)dx + \int_{0}^{\lambda} (1 - F(x))dx \\
&= \lambda(F(b) - 1) + \int_{0}^{\lambda} (1 - F(x))dx,
\end{aligned}$$

e portanto, tomando o limite quando $b \to \infty$,

$$\int_{0}^{\infty} xdF(x) \geq \int_{0}^{\lambda} (1 - F(x))dx.$$

Como isto é válido para todo $\lambda > 0$, tomando o limite quando $\lambda \to \infty$,

$$\int_0^\infty x dF(x) \geq \int_0^\infty (1 - F(x))dx.$$

∎

A desigualdade de Jensen é uma das propriedades da esperança.

Corolário 6.8.7: Desigualdade de Jensen. *Seja φ uma função mensurável e convexa definida na reta. Se X é integrável, então $E(\varphi(X)) \geq \varphi(E(X))$.*

Prova: Pela convexidade de φ, dado algum ponto $(x_0, \varphi(x_0)$ do gráfico de φ, existe uma reta que passa por esse ponto e fica sempre abaixo do gráfico de φ, ou seja, existe algum λ tal que

$$\varphi(x) \geq \varphi(x_0) + \lambda(x - x_0), \forall x.$$

Logo, pela monotonicidade e linearidade da esperança,

$$E\varphi(X) \geq \varphi(x_0) + \lambda(E(X) - x_0).$$

Em particular, para $x_0 = EX$, tem-se $E\varphi(X) \geq \varphi(E(X))$. ∎

O próximo lema estabelece um critério para integrabilidade de variáveis aleatórias.

Lema 6.8.8: *Seja X uma variável aleatória qualquer. Então,*

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E|X| \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n),$$

e, portanto, X é integrável se, e somente se, $\sum_{n=1}^{\infty} P(|X| \geq n) < \infty$.

Prova: Se $x \geq 0$, seja $\lfloor x \rfloor$ a parte inteira de x. Então, a variável aleatória $\lfloor |X| \rfloor$ assume o valor k quando $k \leq |X| < k + 1$ e $0 \leq \lfloor |X| \rfloor \leq |X| \leq \lfloor |X| \rfloor + 1$, então pela monotonicidade e linearidade da esperança,

$$0 \leq E\lfloor |X| \rfloor \leq E|X| \leq 1 + E\lfloor |X| \rfloor.$$

Como $\lfloor |X| \rfloor$ é uma variável aleatória que só assume valores inteiros não-negativos,

$$E\lfloor |X| \rfloor = \sum_{n=1}^{\infty} P(\lfloor |X| \rfloor \geq n) = \sum_{n=1}^{\infty} P(|X| \geq n),$$

logo

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E(|X|) \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n).$$

∎

Se $X^+ = \max(X, 0)$ e $X^- = -\min(X, 0)$, então $X = X^+ - X^-$ e $|X| = X^+ + X^-$. Por definição, $E(X) < \infty$ se, e somente se, $E(X^+) < \infty$ e $E(X^)- < \infty$. Portanto, $E(X) < \infty$ se, e somente se, $E(|X|) < \infty$. De forma análoga, pode-se concluir que $E(\varphi(X)) < \infty$ se, e somente se, $E(|\varphi(X)|) < \infty$ para qualquer função mensurável φ.

O próximo teorema fornece um outro critério para integrabilidade de uma variável aleatória.

Teorema 6.8.9: *Sejam X e Y variáveis aleatórias tais que $Y \geq 0$, Y é integrável e $|X| < Y$. Então, X é integrável.*

Prova: Note que $0 \leq |X| \leq Y$ implica que $0 \leq E(|X|) \leq E(Y)$. Portanto, se $E(Y) < \infty$, então $E(|X|) < \infty$, o que por sua vez implica que $E(X) < \infty$. ∎

Os dois importantes teoremas (Burrill, 1972) a seguir tratam da convergência de esperanças de variáveis aleatórias. O critério de convergência envolvido é o pontual ou seja, $X_n \to X$ se, e somente se, $X_n(w) \to X(w)$ para todo $w \in \Omega$.

Teorema 6.8.10: Teorema da Convergência Monótona. *Sejam $X, X_1, X_2, \ldots$ variáveis aleatórias. Se $0 \leq X_n \uparrow X$, então, $E(X_n) \uparrow E(X)$.*

Teorema 6.8.11: Teorema da Convergência Dominada. *Sejam $Y, X, X_1, X_2, \ldots$ variáveis aleatórias. Considere que Y seja integrável, $|X_n| \leq Y$ e $X_n \to X$. Assim X e X_n são integráveis e $E(X_n) \to E(X)$.*

O próximo exemplo mostra que nem sempre $X_n \to X \Rightarrow E(X_n) \to E(X)$.

Exemplo 6.8.12: Seja $Y \sim U(0,1)$. Considere a seguinte sequência $\{X_1, X_2, \ldots\}$ de variáveis aleatórias: $X_n(\omega) = n$ se $Y(\omega) \in (0, 1/n)$ e $X_n(\omega) = 0$, caso contrário. Então, $X_n(\omega) \to 0, \forall \omega$. Mas, $E(X_n) = 1 \neq 0 = E(0)$, ou seja, $E(X_n) \nrightarrow 0$. ∎

6.9 Exercícios

6.1 Calcule $E(X)$ e $V(X)$ para as seguintes leis de probabilidades:

(a) $P(X=0)=0.2$, $P(X=1)=0.5$, $P(X=2)=0.3$.

(b) $P(X=-1)=0.2$, $P(X=0)=0.3$, $P(X=3)=0.5$.

(c) $P(X=k)=\frac{k}{6}$, para $k=1,2,3$.

(d)

$$f_X(x)=\begin{cases} 2(1-x), & 0<x<1,\\ 0, & \text{caso contrário.}\end{cases}$$

(e)

$$f_X(x)=\begin{cases} \frac{x+2}{18}, & -2<x<4,\\ 0, & \text{caso contrário.}\end{cases}$$

(f)

$$f_X(x)=\begin{cases} \frac{1}{2}(x+1), & -1<x<1,\\ 0, & \text{caso contrário.}\end{cases}$$

(g)

$$f_X(x)=\begin{cases} 6x(1-x), & 0<x<1,\\ 0, & \notin(0,1).\end{cases}$$

(h)

$$f_X(x)=\begin{cases} \frac{2}{x^3}, & 1<x<\infty,\\ 0, & \notin(0,\infty).\end{cases}$$

(i)

$$F_X(x)=\begin{cases} 0, & x<0,\\ \frac{x}{8}, & 0\le x<2,\\ \frac{x^2}{16}, & 2\le x<4,\\ 1, & x\ge 4.\end{cases}$$

6.2 Suponha que (X,Y) seja uma variável aleatória bidimensional tal que

$$f(x,y)=\begin{cases} 4y(1-x), & 0<x<1,\ 0<y<1,\\ 0, & \text{caso contrário.}\end{cases}$$

Determine

(a) $E(XY)$.

(b) $V(X+Y)$.

6.3 Seja X uma variável aleatória tal que

$$F_X(x)=\begin{cases} 0, & x<2,\\ \frac{x-2}{2}, & 2\le x<4,\\ 1, & x\ge 4,\end{cases}$$

e $Y=3X-3$. Determine a média e a variância de Y, 6.4 Usando a desigualdade de Tchebychev estime uma cota superior para a probabilidade de que uma variável aleatória tendo média μ e desvio padrão σ se desvie de μ por menos que 3σ.

6.5 Considere o cabeçote de um disco movendo-se dentro de um cilindro de raio interior a e exterior b. Suponha também que há um número muito grande de cilindros suficientemente próximos para supor continuidade. Sejam X e Y variáveis aleatórias denotando, respectivamente, a posição atual e a posição almejada do cabeçote. Suponha independência de X e Y e que ambas são uniformemente distribuídas no intervalo (a,b). O movimento do cabeçote numa operação de busca é modelado por uma distância dada por $\mid X - Y \mid$. Calcule a distância média percorrida pelo cabeçote numa operação de busca.

6.6 Uma impressora tem uma probabilidade constante de 0.5 de entrar em pane em qualquer um dos 5 dias úteis da semana. Se a máquina não apresentar panes durante a semana, um lucro S é obtido. Se 1 ou 2 panes ocorrerem, é obtido um lucro R ($R < S$). Se 3 ou mais panes ocorrerem, um lucro $-L$ é alcançado. Admita que quando a máquina entra em pane em qualquer um dos dias, ela permanece em pane o resto do dia.

(a) Encontre a distribuição de probabilidade do lucro obtido por semana.

(b) Determine o lucro médio obtido por semana.

6.7 A variância da soma de duas variáveis aleatórias X e Y é dada por

$$V(X+Y) = V(X) + V(Y) + 2Cov(X,Y),$$

onde,

$$Cov(X,Y) = E(XY) - E(X)E(Y),$$

é a covariância entre X e Y.

(a) Calcule $V(X+Y)$ quando $f(x,y) = 6(1-x-y), \;\; 0 < y < 1-x < 1$.

(b) Por que $V(X+Y) = V(X) + V(Y)$ quando X e Y são independentes? Justifique sua resposta.

6.8 Suponha que a variável aleatória bidimensional contínua (X,Y) seja uniformemente distribuída sobre a região limitada pela reta $y = 4$ e pela parábola $y = x^2$.

(a) Estabeleça a distribuição de probabilidade conjunta de (X,Y).

(b) Encontre $f_X(x)$ e $f_Y(y)$.

(c) Mostre que $E(X+Y) = E(X) + E(Y)$.

(d) Verifique se $E(XY) = E(X)E(Y)$.

6.9 Suponha que a variável aleatória bidimensional (X,Y) seja uniformemente distribuída sobre o triângulo de vértices $(-1,3)$, $(0,0)$ e $(1,3)$.

(a) Especifique a distribuição conjunta de (X, Y).

(c) Calcule $V(Y)$ sem calcular $f_Y(y)$.

6.10 Seja

$$f(x,y) = \begin{cases} 2, & 0 < x < y,\ 0 < y < 1 \\ 0, & \text{quaisquer outros casos} \end{cases}$$

a função densidade conjunta do vetor aleatório (X, Y).

Sejam $u(X, Y) = X$, $v(X, Y) = Y$ e $w(X, Y) = XY$.

Mostre que $E(u(X, Y)) \cdot E(v(X, Y)) \neq E(w(X, Y))$. (1.5)

6.11 Suponha que a demanda (procura) por semana de um certo produto seja uma variável aleatória D com distribuição de probabilidade $p_k = P(D = k)$, para $k = 0, 1, 2, \cdots$. Para este produto sabe-se que o preço de custo é C_1, enquanto o preço de venda é C_2. Se o produto não for vendido até o final da semana, deve ser refugado a um custo adicional C_3. Se o fabricante decide fabricar N desses produtos no início da semana, pede-se:

(a) A distribuição de probabilidade da variável aleatória lucro por semana.

(b) O lucro esperado por semana.

6.12 Sejam os inteiros de 1 a 10 e suponha que um deles seja escolhido aleatoriamente. Considere a variável aleatória X como sendo o número de divisores do número sorteado. Calcule o número médio de divisores do número sorteado.

6.13 n mensagens estão sendo enviadas através de um canal de comunicação. Os tempos de duração das mensagens, T_i, $i = 1 \cdots, n$ são aleatórios, e têm a mesma média μ, a mesma variância σ^2 e são independentes.

(a) Encontre a média e a variância do tempo total T de transmissão das n mensagens.

(b) Encontre $T_{\max}$, que é o tempo máximo praticamente possível durante o qual as mensagens podem ser transmitidas. *Sugestão:* $\mu_X \pm 3\sigma_X$, *three sigma rule.*

6.14 Resolva a letra (a) do problema anterior quando os comprimentos das mensagens são dependentes e o coeficiente de correlação entre as variáveis T_i e T_j é r_{ij}.

6.15 A administração de uma rede planeja o momento Y de começo de uma operação como sendo o tempo máximo em que duas operações de suporte, X_1 e X_2, tenham terminado. As variáveis aleatórias X_1 e X_2 são mutuamente independentes e têm densidades, respectivamente, f_{X_1} e f_{X_2}. Encontre a média e a variância da variável Y.

6.16 Uma mensagem enviada através de um canal de comunicação, consiste de n dígitos 0 ou 1, sendo cada um igualmente provável e independentes. Defina uma variável aleatória X como o número de mudanças nos dígitos.

(a) Encontre a média e a variância de X.

(b) Encontre o número máximo praticamente possível de mudanças.

6.17 Se X e Y são variáveis aleatórias independentes, discretas ou contínuas, mostre que, $\forall y \in R_Y$,

$$E(X \mid Y = y) = E(X).$$

6.18 Se (X, Y) tem uma densidade conjunta $f(x, y) = 2$, para $0 < x < y < 1$. Compute:

(a) $E(Y - X)$;

(b) $V(Y - X)$.

6.19 Dada a densidade conjunta do vetor aleatório (X, Y),

$$f(x, y) = 6(1 - x - y), \;\; 0 < y < 1 - x < 1,$$

calcule

(a) as densidades de X e Y;

(b) $E(XY)$.

6.20 Um jogador lança duas moedas não-viciadas. Ganha 1 u.m. (unidade monetária) ou 2 u.m., conforme ocorra uma ou duas caras. Por outro lado, perde 5 u.m. se não ocorrer cara. Ache o valor esperado do jogo e verifique se o mesmo é favorável ao jogador.

Capítulo 7

Principais Variáveis Aleatórias Discretas

Este capítulo descreve os principais modelos de variáveis aleatórias discretas, isto é, as variáveis aleatórias discretas mais comumente encontradas no mundo físico. Dentre essas destacam-se: Bernoulli, Binomial, Poisson, Geométrica, Pascal, Hipergeométrica, Zeta e, como um modelo de uma distribuição discreta multivariada, a Multinomial. Para cada uma delas será dada a distribuição de probabilidade, $P(X = k)$, ou lei de probabilidade[1] esperança, $E(X)$, e variância, $V(X)$.

Uma explicação: parâmetro da distribuição de probabilidade é a entidade sem a qual é impossível calcular probabilidades envolvendo a variável aleatória. O k, em $P(X = k)$, é um dos valores que a variável aleatória assume com probabilidade diferente de zero, isto é, é um dos valores do seu contradomínio. Se o ou os parâmetros da distribuição de probabilidade não são conhecidos, o que acontece em problemas práticos, a Estatística fornece método para estimá-los.

Observação: Muitos cálculos ficam indicados, ficando para o leitor a tarefa de realizá-los, mas, quando for necessário o uso de decimais nos exemplos resolvidos será adotado o sistema de ponto-flutuante $F(10, 3, e_1, e_2)$ e o arredondamento simétrico[2] (Forsythe (1977), Goldberg (1991), Santos e Silva (2010), Sterbenz (1974), Vandergraft (1983), Wilkson (1963)).

7.1 Bernoulli de parâmetro p: $B(p)$

A modelagem de uma situação do mundo físico por uma Bernoulli envolve definir um evento de interesse, A por exemplo, e a ele associar uma probabilidade $p = P(A)$. Portanto, nesta modelagem, o mundo real é dicotômico, isto é, ou A acontece, ou não, neste último caso, acontece seu complementar. Assim, uma Bernoulli pode ser adequada para modelar: o estado de uma impressora, se funcionando ou não; em uma palavra de máquina um dado *bit* ser 1 ou 0. Além de modelar o mundo real, a Bernoulli é básica em desenvolvimentos teóricos como, para conjuntamente com a desigualdade de Tchebychev, provar a Lei dos Grandes Números.

[1]Portanto, $0 \leq P(X = k) \leq 1$, $\forall k$ e $\sum_k P(X = k) = 1$.

[2]Vide Apêndice A

(i) Distribuição de probabilidade.

$$P(X = k) = \begin{cases} q, & k = 0, \\ p, & k = 1, \end{cases}$$

onde $q = 1 - p$ e portanto $\sum_{k=0}^{1} P(X = k) = 1$.

(ii) Esperança.

$$E(X) = 0 \times q + 1 \times p = p.$$

(iii) Variância.

$$E(X^2) = 0^2 \times p + 1^2 \times p = p,$$

logo

$$V(X) = E(X^2) - E(X)^2 = p - p^2 = pq.$$

Exemplo 7.1.1: Suponha uma única posição de memória num computador digital. O objetivo é modelar a distribuição da variável aleatória, X, que conta o número de uns (1's) nesta posição de memória. Imagine que não haja razão para supor que o aparecimento de 0 ou 1 tenham probabilidades diferentes, portanto, assume-se que $P(0) = P(1) = \frac{1}{2}$. Então, $P(X = 0) = P(X = 1) = \frac{1}{2}$. ■

7.2 Binomial de parâmetros n e p: $B(n, p)$

Uma variável binomial conta o número de ocorrências (ou o número de sucessos) de um dado evento A em n realizações de experimentos independentes de Bernoulli, onde $P(A) = p$ permanece constante em todo o desenvolvimento do experimento. Assim, uma binomial é adequada para modelar, entre outros, o número de zeros em uma palavra de máquina de precisão simples ou dupla, o número de processadores em funcionamento em um sistema multiprocessador ou o número de servidores ativos em um dado sistema de computação.

(i) Distribuição de probabilidade.

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

Note que, usando o *teorema binomial*[3] tem-se que

$$\sum_{k=0}^{n} P(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1.$$

[3] $(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$.

(ii) Esperança.

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} k\binom{n}{k}p^k q^{n-k} = \sum_{k=0}^{n} k\frac{n!}{k!(n-k)!}p^k q^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!}p^k q^{n-k} = \sum_{k=1}^{n} n\frac{(n-1)!}{(k-1)!(n-k)!}p^k q^{n-k} \\
&= np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}q^{n-k} = np.
\end{aligned}$$

(iii) Variância.

Um cálculo similar ao de $E(X)$ mostra que

$$E(X^2) = npq + n^2p^2$$

e portanto,

$$V(X) = npq + n^2p^2 - n^2p^2 = npq.$$

Uma variável aleatória relacionada com uma $X \sim B(n,p)$ é $Y = n - X$. Neste caso, Y conta o número de falhas[4] nas n repetições independentes do experimento, sendo então uma $B(n,q)$.

Exemplo 7.2.1: Um sistema de computação on-line tem 20 linhas de comunicação que operam independentemente. A probabilidade que qualquer linha esteja ocupada é 0.6. Qual é a probabilidade que 10 ou mais linhas estejam em operação?

Solução: $A = \{\text{a linha está ocupada}\}$. $P(A) = p = 0.6 \Rightarrow q = 1 - p = 0.4$. Seja $X=$ número de linhas ocupadas, ou em operação. Logo,

$$P(X = k) = \binom{20}{k}0.6^k 0.4^{20-k},\ k = 0, \ldots, 20.$$

Portanto,

$$P(X \geq 10) = \sum_{k=10}^{20}\binom{20}{k}0.6^k 0.4^{20-k}.$$

■

7.3 Poisson de parâmetro λ: $Poisson(\lambda)$

A função de probabilidade Poisson é usualmente utilizada para modelar a contagem do número de ocorrências de eventos aleatórios em um certo tempo t, como por exemplo o número de fótons emitidos por uma fonte de luz de intensidade I fótons/seg em t segundos ($\lambda = It$), o número de clientes chegando em uma fila no tempo t ($\lambda = Ct$), o número de ocorrências de eventos raros no tempo t ($\lambda = Ct$), entre outros.

[4]Na verdade a definição do que sucesso ou falha depende de como a modelagem está sendo realizada.

(i) Distribuição de probabilidade.

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, k \in \{0, 1, \ldots\}.$$

Usando o resultado da expansão em série de Taylor da função exponencial, sabe-se que para todo x real,

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Portanto,

$$\sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1.$$

(ii) Esperança.

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k\frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-1)!} = \sum_{s=0}^{\infty} \frac{e^{-\lambda}\lambda^{s+1}}{s!} \\ &= \lambda e^{-\lambda} \sum_{s=0}^{\infty} \frac{e^{-\lambda}\lambda^s}{s!} = \lambda e^{-\lambda}e^{\lambda} = \lambda. \end{aligned}$$

No cálculo anterior, $k - 1 = s \Rightarrow k = s + 1$.

(iii) Variância.

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2\frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} k\frac{e^{-\lambda}\lambda^k}{(k-1)!} = \sum_{s=0}^{\infty}(s+1)\frac{e^{-\lambda}\lambda^{s+1}}{s!} \\ &= \sum_{s=0}^{\infty} s\frac{e^{-\lambda}\lambda^{s+1}}{s!} + \sum_{s=0}^{\infty} \frac{e^{-\lambda}\lambda^{s+1}}{s!} \\ &= \lambda\sum_{s=0}^{\infty} s\frac{e^{-\lambda}\lambda^s}{s!} + \lambda\sum_{s=0}^{\infty} \frac{e^{-\lambda}\lambda^s}{s!} \\ &= \lambda^2 + \lambda. \end{aligned}$$

Portanto,

$$V(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Exemplo 7.3.1: Se a probabilidade de 0 fótons serem emitidos no tempo $[0, t)$ é igual a 0.1, então qual a probabilidade de que pelo menos 2 fótons serem emitidos no mesmo tempo t?

Solução: $A = \{$emissão de fótons no tempo $t\}$. Seja $X=$ número de fótons serem emitidos no tempo t. Neste problema não foi dado o parâmetro λ, mas foi fornecida uma condição para encontrá-lo.

$$P(X=0) = 0.1 \Rightarrow \frac{e^{-\lambda}\lambda^0}{0!} = 0.1 \Rightarrow e^{-\lambda} = 0.1 \Rightarrow \lambda = 2.302.$$

Portanto, $P(X \geq 2) = 1 - P(X \leq 1) = 1 - e^{-2.302} - 2.302e^{-2.302}$. ■

Exemplo 7.3.2: Um valor mais provável de uma distribuição de Poisson é k^* se

$$P(X = k^*) \geq P(X = k^* + 1)$$

e

$$P(X = k^*) \geq P(X = k^* - 1).$$

Solução: Realizando estes cálculos, estas condições são equivalentes a,

$$k^* \leq \lambda \leq k^* + 1$$

ou

$$\lambda - 1 \leq k^* \leq \lambda.$$

Se k^* é o maior inteiro menor ou igual a λ esta restrição é satisfeita, e portanto este é um valor mais provável desta distribuição. Em outras palavras, k^* é o valor de k que torna máxima a probabilidade na Poisson. ■

7.3.1 Poisson como um Limite de Eventos Raros de Binomial

A distribuição de Poisson pode ser encontrada pelo limite de uma $B(n, p)$, quando n vai para infinito (isto é, o experimento é realizado um número grande de vezes), p é muito pequeno mas np, que é média da binomial, permanece constante. A explanação a seguir[5] motiva como essa aproximação pode ser realizada.

Suponha que chamadas telefônicas cheguem em uma grande central e que em um período particular de três horas (180 minutos) um total de 270 chamadas tenham sido recebidas, ou seja, 1.5 chamadas por minuto. O objetivo é calcular a probabilidade de serem recebidas k chamadas durante os próximos três minutos. É natural pensar que a qualquer instante pode ocorrer uma chamada, portanto a modelagem do problema exige que aproximações sejam feitas.

Para começar, pode-se dividir o intervalo de 3 minutos em nove intervalos de 20 segundos cada um e tratar cada um desses nove intervalos como um ensaio de Bernoulli, durante o qual observa-se uma chamada (sucesso) ou nenhuma chamada (falha), com probabilidade de sucesso igual a $p = 1.5 \times \frac{20}{60} = 0.5$. Desse modo, a tentação é grande para afirmar que a probabilidade de 2 chamadas é igual a $\binom{9}{2}(0.5)^9 = \frac{9}{128}$. Porém, este cálculo ignora a possibilidade de que mais de uma chamada possa ocorrer em um único intervalo. Então, por

[5] de P. Meyer (1983) pág. 187

que não aumentar o número n de subintervalos de tempo de modo que cada subintervalo corresponda a $\frac{180}{n}$ segundos e portanto a probabilidade de ocorrência de uma chamada em um subintervalo seja igual a $p = 1.5 \times \frac{180}{60n}$? Desta maneira $np = 4.5$ permanece constante quando o número de subintervalos cresce. Utilizando novamente o modelo binomial, a probabilidade de ocorrerem k chamadas é dada por: $\binom{n}{k}(\frac{4.5}{n})^k(1-\frac{4.5}{n})^{n-k}$. O que acontece com esta probabilidade quando $n \to \infty$? A resposta, como será visto a seguir, é que esta distribuição tende a uma distribuição de Poisson, sendo este resultado conhecido como *limite de eventos raros*.

Seja a expressão geral da probabilidade binomial,

$$P(X=k) = \binom{n}{k} p^k(1-p)^{n-k} = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!}p^k(1-p)^{n-k}.$$

Como o objetivo é estudar o caso em que np é constante, seja $np = \lambda$, ou seja, $p = \frac{\lambda}{n}$ e $1 - p = \frac{n-\lambda}{n}$. Então,

$$\begin{aligned} P(X=k) &= \frac{n(n-1)\cdots(n-k+1)}{k!}(\frac{\lambda}{n})^k(\frac{n-\lambda}{n})^{n-k} \\ &= \frac{\lambda^k}{k!}((1)(1-\frac{1}{n})\cdots(1-\frac{k-1}{n})(1-\frac{\lambda}{n})^{n-k}. \end{aligned}$$

Fazendo $n \to \infty$, os termos da forma $(1-\frac{j}{n})$, para $1 \leq j \leq k-1$, tendem para 1 e como existe um número fixo $k-1$ deles, o seu produto também tende a 1. O mesmo ocorre com $(1-\frac{\lambda}{n})^{-k}$. Finalmente, por definição do número e, tem-se que $(1-\frac{\lambda}{n})^n \to e^{-\lambda}$ quando $n \to \infty$. Portanto,

$$\lim_{n\to\infty, p\to 0, \lambda=np} P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!},$$

ou seja, obteve-se a expressão de Poisson.

Exemplo 7.3.3: Ao formar números binários com n dígitos, a probabilidade de que um dígito incorreto possa aparecer é 0.002. Se os erros forem independentes, qual é a probabilidade de encontrar k dígitos incorretos em um número binário de 25 dígitos? Se um computador forma 10^6 desses números de 25 dígitos por segundo, qual é a probabilidade de que pelo menos um número incorreto seja formado durante qualquer período de 1 segundo?
Solução: A probabilidade de que k dígitos sejam incorretos em um número binários de 25 dígitos é igual a $\binom{25}{k}(0.002)^k(0.998)^{25-k}$. Em particular, a probabilidade de que pelo menos um dígito seja incorreto é igual a $1-(0.998)^{25} \approx 0.049$. Usando a aproximação pela Poisson então tem-se uma Poisson com parâmetro $25 \times 0.002 = 0.05$, logo a probabilidade de pelos menos um dígito incorreto neste número de 25 dígitos é $1 - e^{-0.05} \approx 0.049$.

A probabilidade de que pelo menos um número incorreto seja formado durante um período de 1 segundo é igual a $1-(0.049)^{10^6} \approx 1 - e^{-49000} \approx 1$. ■

A distribuição de Poisson também pode ser deduzida de específicas hipóteses relativas a determinada classe de fenômenos estudados sendo, neste caso, o parâmetro λ obtido proporcionalmente[6].

[6]Ver o processo de Poisson, Capítulo 10.

Exemplo 7.3.4: Suponha que o número de clientes que chegam em um banco segue uma distribuição de Poisson. Se a probabilidade de chegarem 3 clientes for o triplo da de chegarem 4 clientes em um dado período de 10 minutos. Determine:

(a) Qual o número esperado de clientes que chegam em um período de 1 hora neste banco?

(b) Qual o número mais provável de clientes que chegam em um período de 1 hora neste banco?

Solução: Seja X_{10} = número de clientes que chegam ao banco em 10 minutos.

$$P(X_{10} = 3) = 3P(X_{10} = 4) \Rightarrow \frac{e^{-\lambda}\lambda^3}{3!} = 3 \times \frac{e^{-\lambda}\lambda^4}{4!} \Rightarrow \lambda = 1.333.$$

(a) Seja X_{60} = número de clientes que chegam ao banco em 60 minutos. Para $t = 10$ minutos tem-se que $\lambda = 1.333$, logo para $t = 60$ minutos, $\lambda = \frac{60\times 1.333}{10} = 7.998$ e portanto $E(X_{60}) = 7.998 \approx 8$.

(b) $6.998 \leq \lambda^* \leq 7.998 \Rightarrow \lambda^* = 7$ ou $\lambda^* = 8$. O valor de λ^* dependerá da regra de decisão associada ao problema.

■

7.4 Geométrica de parâmetro p: $G(p)$

A geométrica pode ser utilizada para modelar o número de repetições independentes do lançamento de uma moeda até a primeira ocorrência de cara, tempo de espera medido em unidades de tempo inteiras até a chegada do próximo consumidor em uma fila, ou até a próxima emissão de um fóton. Esta variável, assim como as anteriores, também é uma variável de contagem, só que ela está relacionada à *primeira* ocorrência de sucesso do evento A de interesse na modelagem. Por exemplo, se o evento de interesse é a ocorrência do primeiro 1 numa *string* de zeros e uns, se a variável assumir o valor 10, a string observada foi 0000000001.

(i) Distribuição de probabilidade.

$$P(X = k) = q^{k-1}p, k \in \{1, 2, 3, \ldots\},$$

porque $\{X = k\}$ é equivalente $\underbrace{\overline{A} \cap \overline{A} \cap \ldots \cap \overline{A}}_{k-1} \cap A$.

Utilizando o resultado de uma soma infinita de uma progressão geométrica ilimitada de razão $\mid r \mid< 1$,

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

Logo, esta é uma legítima função probabilidade de massa.

(ii) Esperança.

$$\begin{aligned}
E(X) &= \sum_{k=1}^{\infty} kP(X=k) = \sum_{k=1}^{\infty} kq^{k-1}p \\
&= p\sum_{k=1}^{\infty} kq^{k-1} = p\sum_{k=1}^{\infty} \frac{\mathrm{d}}{\mathrm{d}q}q^k \\
&= p\frac{\mathrm{d}}{\mathrm{d}q}\sum_{k=1}^{\infty} q^k = p\frac{\mathrm{d}}{\mathrm{d}q}(\frac{q}{1-q}) = \frac{1}{p}.
\end{aligned}$$

(iii) Variância. Usando a função geratriz de momentos (Meyer, 1983) tem-se que

$$E(X^2) = \frac{1+q}{p^2}.$$

Logo,

$$V(X) = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$$

Exemplo 7.4.1: Suponha que joga-se uma moeda independentemente até que uma coroa ocorra. Sabe-se que probabilidade de cara igual a $0 < p < 1$. Seja X o número de repetições necessárias até que coroa apareça pela primeira vez na sequência. Qual é a probabilidade do evento $\{X = k\}$ para $k \in \{1, 2, 3, \ldots\}$? Note que para que $X = k$ é necessário que os primeiros $k-1$ lançamentos sejam caras e o k-ésimo lançamento seja coroa, logo, pela independência dos lançamentos, $P(X = k) = p^{k-1}q$. Ou seja, X é uma variável geométrica de parâmetro q. ■

Exemplo 7.4.2: Suponha que X tenha uma distribuição geométrica com parâmetro β. Mostre que para quaisquer dois inteiros positivos s e t,

$$P(X > s+t|X > s) = P(X > t).$$

Solução:

$$P(X > s+t|X > s) = \frac{P(X > s+t, X > s)}{P(X > s)} = \frac{P(X > s+t)}{P(X > s)}.$$

Mas

$$P(X > s+t) = \sum_{k=s+t+1}^{\infty} (1-\beta)^{k-1}\beta = (1-\beta)^{s+t}.$$

Similarmente, $P(X > s) = (1-\beta)^s$ e $P(X > t) = (1-\beta)^t$. Portanto,

$$P(X > s+t|X > s) = \frac{(1-\beta)^{s+t}}{(1-\beta)^s} = (1-\beta)^t = (X > t).$$

Esta propriedade da distribuição geométrica é conhecida como *falta de memória*[7]. ■

[7]Como será visto no capítulo seguinte, a variável Exponencial também tem essa propriedade.

7.5 Pascal de parâmetros r e p: $Pascal(p, r)$

Esta distribuição pode ser considerada como uma generalização da distribuição geométrica. Suponha que o interesse seja calcular a probabilidade de que um experimento tenha de ser repetido k vezes para que um evento A ocorra r vezes. Seja X o número de repetições necessárias a fim de que um evento A possa ocorrer exatamente r vezes. $X = k$ se, e somente se, A ocorrer na k-ésima repetição e A tiver ocorrido $r - 1$ vezes nas $(k - 1)$ repetições anteriores. Uma possível realização do experimento é

$$B = \underbrace{\overline{A} \cap \ldots \cap \overline{A}}_{k-r} \cap \underbrace{A \ldots \cap A}_{r}.$$

Assumindo independência entre os eventos, a probabilidade acima corresponde a

$$\underbrace{q \times \ldots \times q}_{k-r} \times \underbrace{p \times \ldots \times p}_{r} = q^{k-r} p^r.$$

Mas, quantas realizações distintas do evento B são possíveis? A resposta é $\binom{k-1}{r-1}$. Portanto,

(i) Distribuição de probabilidade.

$$P(X = k) = \binom{k-1}{r-1} p^r q^{k-r},\ k \geq r.$$

Se $r = 1$, tem-se que X tem uma distribuição geométrica com parâmetro p.

(ii) Esperança e Variância.

Para calcular $E(X)$ e $V(X)$ pode-se proceder da seguinte maneira. Seja $Z_1, Z_2, \ldots$ uma sequência de variáveis aleatórias tal que Z_1 é o número de repetições do experimento até a primeira ocorrência de um evento A, Z_i é o número de repetições do experimento entre a $(i - 1)$-ésima até e incluindo a i-ésima ocorrência de A, para $i = 2, 3, \ldots, r$, isto é,

$$\underbrace{\overline{A} \cap \ldots \cap \overline{A}}_{Z_1} \cap A \cap \underbrace{\overline{A} \cap \ldots \cap \overline{A}}_{Z_2} \cap A \ldots \cap \underbrace{\overline{A} \cap \ldots \cap \overline{A}}_{Z_r} \cap A.$$

Então, as variáveis Z_i são independentes, cada uma delas tem uma distribuição geométrica com parâmetro p e $X = Z_1 + Z_2 + \cdots + Z_r$. Logo, X pode ser considerada como uma soma de r geométricas independentes, portanto, usando propriedades da esperança e da variância,

$$E(X) = \frac{r}{p}$$

e

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

Exemplo 7.5.1: Suponha que X tenha distribuição binomial com parâmetros n e p e Y tenha uma distribuição de Pascal com parâmetros r e p. Portanto, $P(X \geq r) = P(Y \leq n)$.

Estas duas distribuições tratam de ensaios de Bernoulli repetidos. A distribuição binomial surge quando se tem um número fixo de ensaios e o interesse é o número de sucessos que venham a ocorrer. A distribuição de Pascal é encontrada quando o número de sucessos é fixo, r, e o que é registrado é o número de ensaios necessários para a obtenção dos r sucessos. ■

Observação. Pascal ou binomial negativa? Jain (1991) (página 492) considera Pascal e binomial negativa distintas. A binomial negativa é definida como sendo *o número de falhas antes de ocorrerem r sucessos*. Portanto, se $k = 4$ e $r = 3$, possíveis realizações do experimento são:

$$\overline{A} \cap \overline{A} \cap \overline{A} \cap \overline{A} \cap A \cap A \cap A,$$

$$\overline{A} \cap A \cap \overline{A} \cap A \cap \overline{A} \cap \overline{A} \cap A.$$

Cada um dos eventos acima tem probabilidade p^3q^4. Mas, quantos são? Como a última posição está fixa, tem-se $\binom{4+2}{2}$ arrumações. Portanto,

$$P(X = k) = \binom{k + (r-1)}{r-1} p^r q^k, \; k = 0, 1, \ldots$$

Quando $k = 0$ então ocorreu $\underbrace{A \cap A \cap \ldots \cap A}_{r}$.

Grinstead e Snell (1997), página 186, chamam a Pascal de binomial negativa. Para Meyer (1983), página 204, a distribuição de Pascal pode ser chamada de binomial negativa.

7.6 Hipergeométrica de parâmetros N, D, e n: $H(n, D, n)$

A distribuição hipergeométrica descreve o número de sucessos em uma sequência de n amostras retiradas sem reposição de uma população finita.

Por exemplo, considere que tem-se uma carga com N objetos dos quais D são defeituosos. A distribuição hipergeométrica descreve a probabilidade de que em uma amostra de n objetos distintos escolhidos da carga aleatoriamente exatamente k objetos sejam defeituosos.

(i) Distribuição de probabilidade.

$$P(X = k) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}.$$

Esta probabilidade é positiva se: $N - D \geq n - k$, ou seja $k \geq \max(0, D + n - N)$ e $k \leq \min(n, D)$.

Esta fórmula pode ser entendida assim: existem $\binom{N}{n}$ possíveis amostras sem reposição, $\binom{D}{k}$ maneiras de escolher k objetos defeituosos e $\binom{N-D}{n-k}$ maneiras de preencher o resto da amostra com objetos sem defeito.

Quando a população é grande quando comparada ao tamanho da amostra (ou seja, N for muito maior que n) a distribuição hipergeométrica é aproximada por uma distribuição binomial com parâmetros n (tamanho da amostra) e $p = D/N$ (probabilidade de sucesso em um único ensaio) (Meyer, 1983).

(ii) Esperança.

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} k\frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} = \sum_{k=1}^{n} \frac{D!(N-D)!(N-n)!n!}{k!(D-k)!(n-k)!(N-D-n+k)!N!} \\
&= \frac{nD}{N}\sum_{k=1}^{n} \frac{(D-1)!(N-D)!(N-n)!(n-1)!}{(k-1)!(D-k)!(n-k)!(N-D-n+k)!(N-1)!} \\
&= \frac{nD}{N}\sum_{k=1}^{n} \frac{\binom{D-1}{k-1}\binom{N-D}{n-k}}{\binom{N-1}{n-1}}
\end{aligned}$$

Substituindo no somatório $D^* = D-1, k^* = k-1, n^* = n-1$ e $N^* = N-1$,

$$E(X) = \frac{nD}{N}\sum_{k^*=0}^{n^*} \frac{\binom{D^*}{k^*}\binom{N^*-D^*}{n^*-k^*}}{\binom{N^*}{n^*}} = \frac{nD}{N}.$$

Se $p = \frac{D}{N}$ então, $E(X) = np$.

O somatório acima é igual a soma da função probabilidade de massa de uma variável aleatória hipergeométrica para todos os valores que tem probabilidade positiva, e portanto, é igual a 1.

(iii) Variância.

$$
\begin{aligned}
E(X^2) &= \sum_{k=0}^{n} k^2 \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}} \\
&= \sum_{k=0}^{n} k^2 \times \frac{D}{k} \times \frac{n}{N} \times \frac{\binom{D-1}{k-1}\binom{N-D}{n-k}}{\binom{N-1}{n-1}} \\
&= D \times \frac{n}{N} \sum_{k=0}^{n} k \frac{\binom{D-1}{k-1}\binom{N-D}{n-k}}{\binom{N-1}{n-1}} \\
&= D \times \frac{n}{N} \sum_{k=0}^{n} (k+1) \frac{\binom{D-1}{k}\binom{N-D}{n-k-1}}{\binom{N-1}{n-1}} \\
&= D \times \frac{n}{N} (\sum_{k=0}^{n} k \times \frac{D-1}{k} \frac{n-1}{N-1} \frac{\binom{D-2}{k-1}\binom{N-D}{n-k-1}}{\binom{N-2}{n-2}} + \sum_{k=0}^{n} \frac{\binom{D-1}{k}\binom{N-D}{n-k-1}}{\binom{N-1}{n-1}}) \\
&= D \times \frac{n}{N} (\frac{(D-1)(n-1)}{N-1} \sum_{k=0}^{n} \frac{\binom{D-2}{k-1}\binom{N-D}{n-k-1}}{\binom{N-2}{n-2}} + 1) \\
&= D \times \frac{n}{N} (\frac{(D-1)(n-1)}{N-1} + 1).
\end{aligned}
$$

Logo,

$$
\begin{aligned}
V(X) &= D \times \frac{n}{N} (\frac{(D-1)(n-1)}{N-1} + 1) - D^2 \times \frac{n^2}{N^2} \\
&= D \times \frac{n}{N} (\frac{(D-1)(n-1)}{N-1} + 1 - D \times \frac{n}{N}) \\
&= n \times \frac{D}{N} \times \frac{N-n}{N-1} (1 - \frac{D}{N}).
\end{aligned}
$$

Fazendo a mesma substituição no caso de $E(X)$ tem-se que:

$$
V(X) = np(1-p)\frac{N-n}{N-1} = npq\frac{N-n}{N-1}.
$$

Exemplo 7.6.1: Suponha que uma urna contém 20 bolas brancas e 10 bolas pretas. Se 4 bolas são retiradas da urna. Determine:

(a) A probabilidade de pelo menos uma bola seja branca, se as bolas são retiradas com reposição.

(b) A probabilidade de pelo menos uma bola seja branca, se as bolas são retiradas sucessivamente e sem reposição.

(c) A probabilidade de pelo menos uma bola seja branca, se as bolas são retiradas simultaneamente.

Solução: Seja $B_i = \{i\text{-ésima bola é branca}, i = 1, 2, 3, 4.\}$

(a)

$$\begin{aligned} P(B_1 \cup \cdots \cup B_4) &= 1 - P(\overline{B_1 \cup \cdots \cup B_4}) \\ &= 1 - P(\overline{B_1} \cap \cdots \cap \overline{B_4}) \\ &= 1 - P(\overline{B_1}) \times \cdots \times P(\overline{B_4}) \\ &= 1 - (\frac{10}{30})^4. \end{aligned}$$

Este item também poderia ser resilvido usando o Teorema da Inclusão-Exclusão, mas daria muito trabalho. Observe que foi usado o fato dos eventos serem independentes.

(b)

$$\begin{aligned} P(B_1 \cup \cdots \cup B_4) &= 1 - P(\overline{B_1 \cup \cdots \cup B_4}) \\ &= 1 - P(\overline{B_1} \cap \cdots \cap \overline{B_4}) \\ &= 1 - P(\overline{B_1})P(\overline{B_2}|\overline{B_1}) \times \cdots \times P(\overline{B_4}|\overline{B_1} \cap \cdots \cap \overline{B_4}) \\ &= 1 - \frac{10}{30}\frac{9}{29}\frac{8}{28}\frac{7}{27}. \end{aligned}$$

(c) Seja $X = \{\text{número de bolas brancas retiradas da urna.}\}$ Logo

$$P(X = k) = \frac{\binom{20}{k}\binom{10}{4-k}}{\binom{30}{4}}, \; k = 0, \ldots, 4.$$

Portanto, $P(X \geq 1) = \sum_{k=1}^{4} \frac{\binom{20}{k}\binom{10}{4-k}}{\binom{30}{4}}$.

■

Exemplo 7.6.2: Por engano 3 peças defeituosas foram misturadas com boas formando um lote com 12 peças no total. Escolhendo ao acaso 4 dessas peças, determine a probabilidade de encontrar:

(a) Pelo menos 2 defeituosas.

(b) No máximo 1 defeituosa.

(c) No mínimo 1 boa.

Solução: Seja $X = \{\text{número de peças defeituosas no lote.}\}$ Logo

$$P(X = k) = \frac{\binom{3}{k}\binom{9}{4-k}}{\binom{12}{4}}, \; k = 0, \ldots, 4.$$

(a) $P(X \geq 2) = P(X = 2) + P(X = 3) = \frac{\binom{3}{2}\binom{9}{2}}{\binom{12}{4}} + \frac{\binom{3}{3}\binom{9}{1}}{\binom{12}{4}}$.

(b) $P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{\binom{3}{0}\binom{9}{4}}{\binom{12}{4}} + \frac{\binom{3}{1}\binom{9}{3}}{\binom{12}{4}}$.

(c) Seja $Y = \{$número de peças perfeitas na amostra.$\}$. O contradomínio de Y é $R_Y = \{1, 2, 3, 4\}$ e portanto $P(Y \geq 1) = 1$.

■

7.7 Zeta ou Zipf de parâmetro $\alpha > 1$: $Z(\alpha)$

A função probabilidade Zeta ou Zipf é um exemplo de uma distribuição de caudas-pesadas[8] cuja importância cresceu desde meados dos anos 1990. As aplicações desta função de probabilidade incluem: número de consumidores afetados por um *blackout*, tamanhos de arquivos solicitados em transferência via *Web* e atraso de pacotes na *internet*.

(i) Distribuição de probabilidade.

$$P(X = k) = \frac{k^{-\alpha}}{\zeta(\alpha)}, k = 1, 2, \ldots, \ \alpha > 1$$

onde $\zeta(\alpha) = \sum_{j=1}^{\infty} j^{-\alpha}$ é conhecida como a função Zeta de Riemann, sendo convergente para $\alpha > 1$.

(ii) Esperança.

$$E(X) = \sum_{k=1}^{\infty} k \frac{k^{-\alpha}}{\zeta(\alpha)} = \frac{1}{\zeta(\alpha)} \sum_{k=1}^{\infty} k^{-(\alpha-1)} = \frac{\zeta(\alpha-1)}{\zeta(\alpha)}, \ \alpha > 1.$$

(iii) Variância.

$$E(X^2) = \sum_{k=1}^{\infty} k^2 \frac{k^{-\alpha}}{\zeta(\alpha)} = \frac{1}{\zeta(\alpha)} \sum_{k=1}^{\infty} k^{(-\alpha-2)} = \frac{1}{\zeta(\alpha)} \zeta(\alpha - 2), \alpha > 2.$$

Logo,

$$V(X) = \frac{1}{\zeta(\alpha)} \zeta(\alpha - 2) - (\frac{\zeta(\alpha-1)}{\zeta(\alpha)})^2.$$

Exemplo 7.7.1: Os tamanhos de arquivos armazenados em um grande sistema de arquivos Unix seguem uma distribuição Zeta com parâmetro α quando estes tamanhos são medidos em kilobytes[9].

(a) Se os tamanhos dos arquivos de 1KB são 10000 vezes mais prováveis que tamanhos de arquivos de 1MB, então qual o valor do parâmetro α?

[8] Se uma variável aleatória tem caudas-pesadas então tem probabilidades grandes nas caudas.

[9] 1024bytes = 1kilobyte = 1KB; 1024kilobytes = 1megabyte = 1MB; 1024megabytes = 1gigabytes = 1GB; 1024gigabytes = 1terabyte = 1TB; 1024terabytes = 1petabyte = 1PB; 1024petabytes = 1exabyte = 1EB; 1024exabytes = 1zettabyte = 1ZB; 1024zettabytes = 1yottabyte = 1YB.

(b) Quanto mais provável são tamanhos de arquivos de 1MB em comparação com tamanhos de arquivos de 1GB?

Solução:

(a) Seja $X = \{\text{tamanho dos arquivos em KB}\}$. Logo,

$$P(X = 1) = 10000P(X = 1024) \Rightarrow \frac{1}{\zeta(\alpha)} = \frac{1024^{\alpha}}{\zeta(\alpha)} \Rightarrow \alpha = 1.329.$$

(b) $P(X = 1024) = \frac{1024^{-\alpha}}{\zeta(\alpha)}$ e $P(X = 1024^2) = \frac{(1024^2)^{-\alpha}}{\zeta(\alpha)}$. Portanto, $P(X = 1024^2) = 1024^{-\alpha}P(X = 1024)$, sendo pouquiíssimo provável pois $1024^{-1.329} \approx 9.984 \times 10^{-5}$.

■

7.8 Multinomial

A Multinomial é uma distribuição conjunta de variáveis aleatórias discretas, que pode ser considerada como uma generalização da distribuição binomial.

Considere um experimento aleatório qualquer e suponha que o espaço amostral deste experimento é particionado em k eventos $\{A_1, A_2, \ldots, A_k\}$, onde o evento A_i tem probabilidade p_i. Suponha que se repita este experimento n vezes de maneira independente e seja X_i o número de vezes que o evento A_i ocorreu nestas n repetições. Então,

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) = \frac{n!}{n_1!n_2!\cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k},$$

onde $\sum_{i=1}^{k} n_i = n$.

Lembrando que o número de maneiras de arranjar n objetos, n_1 dos quais é de uma espécie, n_2 dos quais é de uma segunda espécie, $\ldots$, n_k dos quais são de uma k-ésima espécie é dado pelo coeficiente multinomial $\frac{n!}{n_1!n_2!\cdots n_k!}$, que corresponde a uma permutação com repetição.

Exemplo 7.8.1: Uma impressora está apresentando uma média de 2 erros por página impressa. Sabe-se que não há razão para supor que exista qualquer relação entre os erros (ou acertos) de uma página para outra, e que o fenômeno obedece a uma lei de Poisson.

(a) Qual é a probabilidade de que uma página não apresente erros?

(b) Uma amostra aleatória de 10 páginas é selecionada de todas as páginas impressas pela impressora. Qual é a probabilidade de que essa amostra apresente 2 ou mais páginas sem erro?

(c) Qual é a probabilidade de que a primeira página sem erro apareça na décima impressão?

(d) Suponha que páginas são impressas até que ocorram 3 sem erro. Qual é a probabilidade de que seja necessário imprimir 5 páginas para que isto aconteça?

(e) Cada página impressa é classificada como *excelente* se não apresentar erros, *boa*, se contiver de 1 a 2 erros e *inaceitável*, se contiver 3 ou mais erros. Se 15 páginas são inspecionadas, qual é a probabilidade de que 8, 5 e 2 páginas, respectivamente, pertencam a cada uma das categorias acima?

Solução:

(a) Seja $N_p = \{$número de erros por página.$\}$. Logo,

$$N_p \sim Poisson(\lambda = 2)$$

e a probabilidade pedida é

$$P(N_p = 0) = e^{-2} = 0.136.$$

(b) Seja $N = \{$número de páginas sem erro.$\}$. Logo,

$$N \sim B(n = 10, p = P(N_p = 0) = 0.136).$$

Portanto,

$$P(N \geq 2) = 1 - P(N \leq 1) = 1 - \binom{10}{0} p^0 q^{10} - \binom{10}{1} p^1 q^9.$$

(c) $N_1 = \{$número de páginas impressas até aparecer a primeira sem erro.$\}$

$$N_1 \sim G(p = P(N_p = 0))$$

e assim,

$$P(N_1 = 10) = (1 - e^{-2})^9 e^{-2}.$$

(d) $N_2 = \{$número de páginas que tem de ser impressas para aparecerem 3 sem erro.$\}$,

$$N_2 \sim Pascal(p = P(N_p = 0)).$$

Logo,

$$P(N_2 = 5) = \binom{5-1}{3-1} (e^{-2})^3 (1 - e^{-2})^{5-3}.$$

(e) A distribuição aqui é uma Multinomial. Seja X_i, $i = 1, 2, 3$, número de páginas classificadas como *excelente*, *boa* e *inaceitável*, respectivamente. Portanto,

$$P(X_1 = 8, X_2 = 5, X_3 = 2) = \frac{15!}{8!5!2!} p_1^8 p_2^5 p_3^2,$$

onde

$$p_1 = P(N_p = 0),\ p_2 = P(1 \leq N_p \leq 2),\ p_3 = P(N_p \geq 3).$$

■

Exemplo 7.8.2: Em determinado cruzamento da BR101 caminhões transitam exibindo placas de vários estados; estima-se que em um dado intervalo de tempo, t, as probabilidades de transitarem neste cruzamento caminhões com placas do Ceará (CE), Amazonas (AM) e Maranhão (MA), são, respectivamente, $\frac{1}{25}$, $\frac{1}{100}$ e $\frac{1}{80}$, sendo $\frac{15}{16}$ a probabilidade de caminhões com placas que não sejam destes três estados.

(a) Se no intervalo de tempo dado passarem neste cruzamento 9 caminhões, qual é a probabilidade de que, dentre os 9, um igual número deles proceda somente de cada um dos referidos estados?

(b) Sabe-se que o número de caminhões com placa do Ceará, que transitam neste cruzamento e no dado intervalo de tempo, segue uma distribuição de Poisson com taxa (ou média) $\frac{1}{25}$ caminhões por dia. Qual é a probabilidade de que em uma semana, pelo menos um caminhão apresente placa do Ceará?

(c) Sabe-se que 2004 terá 53 semanas. Qual é a probabilidade de que o número de semanas onde trafega pelo menos um caminhão com placa do Ceará seja exatamente 2?

(d) Qual é a probabilidade de que a décima semana seja a primeira onde transita pelo menos um caminhão com placa do Ceará?

(e) Qual é a probabilidade de que seja necessário transcorrerem 20 semanas a fim de que em exatamente 3 delas pelos menos um caminhão apresente placa do Ceará?

Solução: Sejam os eventos $C = \{$ a placa é do Ceará$\}$, $A = \{$ a placa é do Amazonas$\}$, $M = \{$ a placa é do Maranhão$\}$ e $N = \{$ a placa não é do CE, AM e MA$\}$. Portanto, $P(C) = \frac{1}{25}$, $P(A) = \frac{1}{100}$, $P(M) = \frac{1}{80}$ e $P(N) = \frac{15}{16}$.

(a) Distribuição: Multinomial. Seja X_i, $i = 1, 2, 3, 4$, respectivamente, número de caminhões com placas do CE, AM, MA e com placas que não sejam destes três estados.

$$P(X_1 = 3, X_2 = 3, X_3 = 3, X_4 = 0) = \frac{9!}{3!3!3!0!}(\frac{1}{25})^3(\frac{1}{100})^3(\frac{1}{80})^3(\frac{15}{16})^0.$$

(b) $X_1 \sim Poisson(\lambda = \frac{1}{25})$. Portanto, se em 1 dia, $\lambda = \frac{1}{25}$, em uma semana, isto é, 7 dias, $\lambda = \frac{7}{25}$.. Seja $X_s = \{$número de caminhões com placa do Ceará trafegando em uma semana$\}$. Logo,

$$X_s \sim Poisson(\lambda = \frac{7}{25}),$$

e

$$P(X_s \geq 1) = 1 - P(X_s < 1) = 1 - e^{-\frac{7}{25}}.$$

(c) Seja $X_C = \{$número de semanas onde trafega pelo menos um caminhão com placa do Ceará$\}$. Então

$$X_C \sim B(n = 53, p = 1 - e^{-\frac{7}{25}})$$

e

$$P(X_C = 2) = \binom{53}{2} p^2(1-p)^{51}.$$

(d) Distribuição Geométrica.

$$P(X_C = 10) = (e^{-\frac{7}{25}})^9(1 - e^{-\frac{7}{25}}).$$

(e) Distribuição de Pascal.

$$P(X_C = 3) = \binom{19}{2}(1 - e^{-\frac{7}{25}})^3(e^{-\frac{7}{25}})^{17}.$$

■

7.9 Exercícios

7.1 Num canal de transmissão com ruído, são transmitidas independentemente 20 cópias de um mesmo pacote. Seja 0.4 a probabilidade de transmissão com sucesso de qualquer uma das cópias. Considere o número de cópias enviadas com sucesso como sendo a variável aleatória de interesse.

(a) Especifique a distribuição de probabilidade ou função densidade dessa variável aleatória.

(b) Qual a probabilidade de que todas as cópias sejam enviadas com sucesso.

7.2 Uma mensagem, enviada em código binário, consiste de uma sequência de símbolos 0 ou 1 todos com igual probabilidade e independentes uns dos outros. Uma sequência do mesmo símbolo é do tipo $000\cdots 0$ ou 11, etc. Seja uma dessas sequências tomadas aleatoriamente. A variável X é o número de símbolos iguais na sequência. Encontre $P(X \geq k)$.

7.3 Uma fábrica produz 10 recipientes de vidro por dia. Deve-se supor que exista uma probabilidade constante $p = 0.1$ de produzir um recipiente defeituoso. Antes que esses recipientes sejam estocados, eles são inspecionados e os defeituosos são separados. Admita que exista uma probabilidade constante $r = 0.1$ de que um recipiente defeituoso seja mal classificado. Faça X igual ao *número de recipients classificados como defeituosos ao fim de um dia de produção*. (Admita que todos os recipientes fabricados em um dia sejam inspecionados naquele dia e que os recipientes perfeitos sempre são classificados corretamente).

(a) Obtenha a expressão de $P(X = k)$.

(b) Calcule $P(X \geq 2)$.

7.4 Uma rodovia está dividida em 8 trechos de igual comprimento, cada qual sob jurisdição de uma guarnição de polícia rodoviária e todos igualmente perigosos. Sabendo-se que nessa rodovia há, em média, 6 desastres por dia, calcular a probabilidade de que, (a) em determinado dia haja quatro trechos sem desastre, (b) 3 trechos com um desastre cada e (c) um trecho com mais de um desastre.

7.5 Seja uma variável aleatória Binomial de parâmetros n e p, $P(X = k) = \binom{n}{k}p^k(1-p)^{n-k},\ k = 0, \ldots, n$. Qual o valor de p onde $P(X = k)$ atinge um máximo, ou mínimo, quando n é par e k é a metade de n?

7.6 Seja X o número obtido de 1's em um número binário escrito na expansão b-ádica, precisão simples, normalizado (1 dígito para o sinal, 8 para o expoente e 23 para a mantissa (ver Apêndice 1)). Estabeleça a distribuição de probabilidade de X.

7.7 Que é mais provável, quando você compete com uma pessoa tão hábil quanto você:

(a) você ganhar três jogos de quatro ou cinco jogos de oito?

(b) não menos que três jogos de quatro ou pelo menos cinco jogos de oito?

7.8 Um centro de processamento de dados comprou um lote de 5000 *chips*, dos quais 1000 foram manufaturados pela fábrica A e o restante pela B. Sabe-se que 10% dos *chips* produzidos por A e 5% dos produzidos por B, respectivamente, são defeituosos.

(a) Um *chip* é escolhido aleatoriamente do lote. Qual é a probabilidade de que seja defeituoso?

(b) Um *chip* é escolhido aleatoriamente do lote, observado, e constata-se que é defeituoso. Qual é a probabilidade de que tenha sido produzido por A?

(c) Suponha que uma amostra de 20 *chips* seja retirada aleatoriamente do lote comprado. Qual será a probabilidade de se encontrar na amostra pelo menos 1 defeituoso?

7.9 Um número binário de n dígitos é escrito onde cada dígito é 0 ou 1 independentemente uns dos outros. A variável aleatória X é o número de dígitos 1. Encontre a probabilidade dos seguintes eventos: (a) $\{X = m\}$; (b) $\{X \geq m\}$; (c) $\{X < m\}$.

7.10 Considere um número escrito na expansão b-ádica

$$x = *d_n d_{n-1} \dots d_1 d_0 . d_{-1} d_{-2} \dots = \sum_{i=n}^{-\infty} d_i b^i,$$

com $* = +$, $b = 10$ e 3 dígitos na mantissa (ver Apêndice 1). Suponha tambem que $d_i = 0, \cdots, 9$ independentemente e com a mesma probabilidade. A variável aleatória X representa cada um dos possíveis números x. Construa a distribuição de probabilidade de X e encontre seu valor médio.

7.11 Um técnico necessita 4 placas para montar um determinado circuito. Encontra 10 na sua oficina, mas sabe que 4 estão defeituosas. Seleciona então 5 dentre as 10 placas. Encontre a probabilidade de que não menos que 4 dentre as 5 estejam perfeitas.

7.12 Uma variável aleatória Y é uma fração própria com n casas decimais. Cada dígito, independentemente uns dos outros, pode ser 0 ou 1 com probabilidade $1/2$. Construa a distribuição de probabilidade de Y e encontre seu valor médio.

7.13 Uma variável aleatória X tem uma distribuição de Poisson com média 3. Encontre a probabilidade de que

(a) X assuma valores menores que sua média.

(b) X assuma valores positivos.

7.14 Revisadas as provas de um livro, verificou-se que há, em média, 2 erros em cada 5 páginas. Em um livro de 100 páginas, estimar quantas não precisam ser modificadas, por não apresentarem erros.

7.15 O número de mensagens que chegam em uma rede tem uma distribuição geométrica com parâmetro p. Para $p = 0.2$, calcule a média, variância, desvio padrão, o coeficiente de variação e plote a distribuição de probabilidade.

7.16 O número de pedidos de I/O recebidos por um disco durante um dado intervalo de tempo segue uma distribuição de Poisson com parâmetro λ. Para $\lambda = 8$ determine a média, variância, desvio padrão e o coeficiente de variação e plote a distribuição de probabilidade.

7.17 Dois processos (ou duas variáveis aleatórias) independentes de Poisson emergem em um disco[10]. Cada um deles tem, respectivamente, parâmetros λ_x e λ_y. Determine o seguinte:

(a) Média de $x + y$.

(b) Variância de $x + y$.

(c) Média de $x - y$.

(d) Variância de $x - y$.

(e) Média de $3x - 4y$.

(f) Coeficiente de variação de $3x - 4y$.

7.18 Um disco rígido recebe em média 2 pedidos de I/O a cada 17 msec, segundo uma distribuição de Poisson.

(a) Qual a probabilidade de que o número de pedidos seja maior que 1, no mesmo tempo considerado?

(b) Este disco é observado durante 10 intervalos de mesmo tempo acima. Qual é a probabilidade de que em ao menos um dos 10 intervalos de tempo o número de pedidos seja maior que 1?

(c) Qual será a probabilidade de que em 34 msec o número de pedidos seja maior que 1?

7.19 Use a aproximação de Poisson para calcular a probabilidade de que no máximo 2 dentre 50 motoristas tenham perdido pontos na carteira de habilitação, se usualmente 5% os perdem.

7.20 Uma companhia de áerea nacional tem observado que 5% das pessoas que fazem reserva para um determinado vôo desistem. A companhia decide então vender 20 bilhetes para este vôo quando só dispõe de 18 lugares. Qual a probabilidade do avião acomodar todos os passageiros?

[10]O leitor, principalmente se aluno de graduação precisa ver uma notação diferente da que é usada em livros de probabilidade inclusive com respeito à nomenclatura para variáveis aleatórias.

7.21 Foguetes são lançados até que o primeiro lançamento bem sucedido tenha ocorrido. Se isso não ocorrer até 5 tentativas, o experimento é suspenso e o equipamento inspecionado. Admita que exista uma probabilidade constante de 0.8 de haver um lançamento bem sucedido e que os sucessivos lançamentos sejam independentes. Suponha que o custo do primeiro lançamento seja k dólares, enquanto os lançamentos subseqüentes custam $k/3$ dólares. Sempre que ocorre um lançamento bem sucedido, uma certa quantidade de informação é obtida, a qual pode ser expressa como um ganho financeiro de c dólares. Seja T o custo líquido desse experimento.

(a) Estabeleça a distribuição de probabilidade de T.

(b) Determine o custo líquido esperado.

7.22 O computador de uma fábrica, que trabalha ininterruptamente, eventualmente falha. O número de falhas pode ser considerado como tendo uma distribuição de Poisson, com número médio de falhas por dia de 1.5; encontre as probabilidades dos seguintes eventos:

(a) $A = \{$o computador falha pelo menos uma vez durante o dia$\}$,

(a) $B = \{$o computador falha não menos que três vezes durante uma semana$\}$.

7.23 Num processo de fabricação 10% das peças são consideradas defeituosas. As peças são acondicionadas em caixas com 5 unidades cada uma.

(a) Qual a probabilidade de haver exatamente 3 peças defeituosas numa caixa?

(b) Qual a probabilidade de haver duas ou mais peças defeituosas numa caixa?

7.24 Um cubo é formado com chapas de plástico de 10×10cm. Em média aparecem 50 defeitos a cada metro quadrado de plástico, segundo uma distribuição de Poisson.

(a) Qual a probabilidade de uma determinada face apresentar exatamente 2 defeitos?

(b) Qual a probabilidade de que pelos menos 5 faces sejam perfeitas?

(c) Qual a probabilidade de que o cubo apresente no mínimo 2 defeitos?

7.25 Um laboratório tem 15 pc's dos quais 5 estão desligados. Um grupo de 6 estudantes entra na sala e, aleatoriamente, cada um escolhe um pc. Qual a probabilidade de que, entre os escolhidos

(a) exatamente dois estejam desligados?

(b) pelo menos um esteja ligado?

(c) pelo menos dois estejam desligados?

7.26 Um carro só tem 4 semáforos em seu percurso. Em cada um deles, independentemente, a probabilidade do carro parar é p. Seja uma variável aleatória X, definida como sendo o número de semáforos que o carro passa antes de parar pela primeira vez.

(a) Estabeleça a distribuição de probabilidade de X. Prove que a expressão encontrada é realmente uma distribuição de probabilidade.

(b) Calcule o número médio de semáforos nos quais o carro passa antes de parar pela primeira vez, para $p = 1/4$.

7.27 Pacientes chegam a um laboratório médico de acordo com uma distribuição de Poisson, com uma média de 2 pacientes a cada 15 minutos.

(a) Sabendo-se que o laboratório abre às 6:00, determine a probabilidade de que cheguem exatamente 4 pacientes até 6:30.

(b) O laboratório funciona das 6:00 às 20:00, sem interrupção. Seja X o número de horas nesse período em que chegam exatamente 8 pacientes. Estabeleça a distribuição de probabilidade de X.

7.28 Admita que o número de navios que chegam a um porto segue uma distribuição de Poisson de média igual a dois navios por dia.

(a) Qual é a probabilidade de que, em um dia qualquer, cheguem, no máximo, 3 navios?

(b) A chegada de navios a esse porto é observada durante 200 dias. Nesse período, qual é o número esperado de dias em que chega apenas um navio?

7.29 Uma fonte mineral contém um número médio de 4 bactérias por centímetro cúbico. Dez tubos de ensaio são enchidos com esse líquido. Supondo que a distribuição de Poisson é aplicável, encontre a probabilidade de que todos os 10 tubos de ensaio apresentem bactérias.

7.30 Na produção de certo tipo de tecidos, os defeitos de produção ocorrem de acordo com uma distribuição de Poisson com taxa de um defeito por cada 2 m^2.

(a) Qual é a probabilidade de que um corte de 2.5 m^2 tenha um ou mais defeitos?

(b) Qual é a probabilidade de que em três cortes de 2.5 m^2, dois sejam perfeitos e um tenha um único defeito?

7.31 Numa estrada pouco movimentada passam, em média, 2 carros por minuto. Supondo a média estável, calcule a probabilidade de que em 2 minutos passem (a) mais de 1 carro, (b) exatamente 4 carros.

7.32 Um celular recebe em média 2 chamadas por hora. Qual a probabilidade de que em 4 horas receba (a) no máximo 2 chamadas? (b) exatamente 3 chamadas.

7.33 Um sistema de computador tem 10 linhas de comunicação cada uma operando independentemente uma da outra. Sabe-se que a probabilidade de que uma linha esteja em uso é 0.6.

(a) Qual é a probabilidade de que 9 ou mais linhas estejam ocupadas?

(b) O administrador do sistema resolve aumentar o número de linhas para 30. Neste caso, qual é a probabilidade de que 9 ou mais delas estejam ocupadas?

Capítulo 8

Principais Variáveis Aleatórias Contínuas

Neste capítulo serão exploradas as principais variáveis aleatórias contínuas. A distribuição Normal tanto é aplicada em problemas práticos quanto teóricos, sendo básica para o desenvolvimento de outras variáveis aleatórias. As distribuições Exponencial, Log-normal e Weibull são fundamentais em modelagem de desempenho de sistemas e confiabilidade. Distribuições como a t-Student, χ^2 e F são úteis no cálculo de intervalos de confiança e em teste de hipóteses. A distribuição de Pareto aplica-se em fenômenos que apresentam grande variabilidade nas observações, o que implica no aumento da variância.

8.1 Uniforme de parâmetros a e b: $U(a, b)$

(i) Densidade.

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in (a,b),\ a,b \in \mathbb{R} \\ 0, & x \notin (a,b). \end{cases}$$

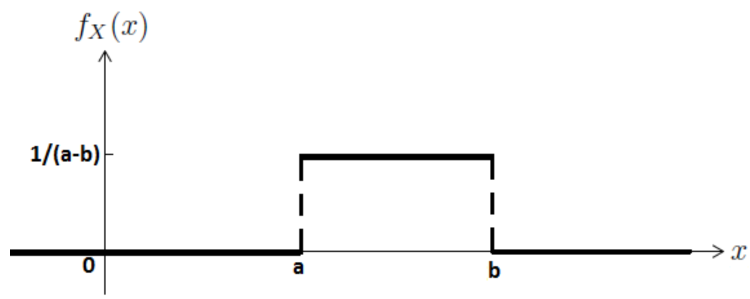

Figura 8.1: Densidade de probabilidade de uma uniforme

(ii) Esperança.

$$E(X) = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}.$$

(iii) Variância.

$$V(X) = E(X^2) - E(X)^2 = \frac{(b-a)^2}{12},$$

onde $E(X)^2 = \int_a^b \frac{x^2}{b-a} dx$.

(iv) Função de distribuição acumulada.

$$F_X(x) = \begin{cases} 0, & x < a, \\ \int_{-\infty}^{x} \frac{1}{b-a} dt = \frac{x-a}{b-a}, & a \leq x < b, \\ 1, & b \leq x < +\infty. \end{cases}$$

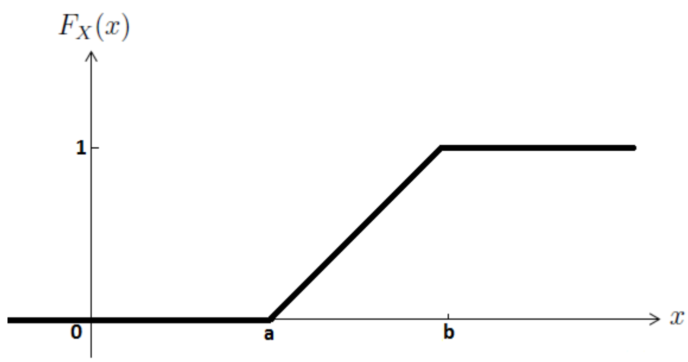

Figura 8.2: Função densidade de probabilidade acumulada de uma uniforme

Exemplo 8.1.1: Um ponto é escolhido ao acaso sobre uma reta de comprimento L. Qual é a probabilidade de que o quociente do segmento mais curto para o mais longo seja menor que $\frac{1}{4}$?

Solução. Seja X a posição do ponto na reta de comprimento L. Dependendo de onde o ponto é escolhido tem-se duas situações:

(a) segmento mais curto X e mais longo $L - X$,

(b) segmento mais curto $L - X$ e mais longo X.

Se (a),

$$\frac{X}{L-X} < \frac{1}{4} \Rightarrow X < \frac{L}{5} \Rightarrow P(X < \frac{L}{5}) = \int_0^{\frac{L}{5}} \frac{1}{L} dx = \frac{1}{5}.$$

Se (b),

$$\frac{L-X}{X} < \frac{1}{4} \Rightarrow X > \frac{4L}{5} \Rightarrow P(X > \frac{4L}{5}) = \int_{\frac{4L}{5}}^{L} \frac{1}{L} dx = \frac{1}{5}.$$

Portanto, a probabilidade pedida é

$$P(X < \frac{L}{5}) + P(X > \frac{4L}{5}) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}.$$

■

8.2 Exponencial de parâmetro λ: $Exp(\lambda)$

A densidade exponencial pode ser utilizada para modelar (i) o tempo de vida de componentes que falham sem efeito de idade, (ii) o tempo de espera entre sucessivas chegadas de fótons, (iii) emissões de elétrons de um cátodo, (iv) chegadas de consumidores, e (v) duração de chamadas telefônicas, entre outros.

(i) Densidade.

$$f_X(x) = \begin{cases} f_X(x) = \lambda e^{-\lambda x}, & x > 0,\ \lambda \in \mathbb{R}^+ \\ 0, & x \leq 0. \end{cases}$$

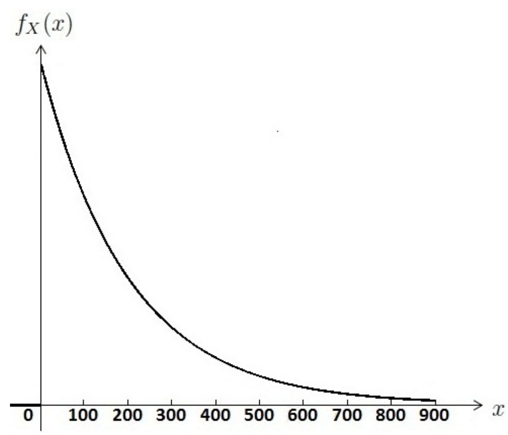

Figura 8.3: Densidade de probabilidade de uma exponencial

(ii) Esperança.

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x} dx = -xe^{-\lambda x}|_0^\infty + \int_0^\infty e^{-\lambda x} dx = \frac{-e^{-\lambda x}}{\lambda}|_0^\infty = \frac{1}{\lambda}.$$

(iii) Variância.

Para o cálculo da variância, inicialmente será calculado o segundo momento.

$$E(X^2) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = -x^2 e^{-\lambda x}|_0^\infty + 2\int_0^\infty xe^{-\lambda x} dx = \frac{2}{\lambda^2}.$$

Portanto,

$$V(X) = E(X)^2 - (E(X))^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

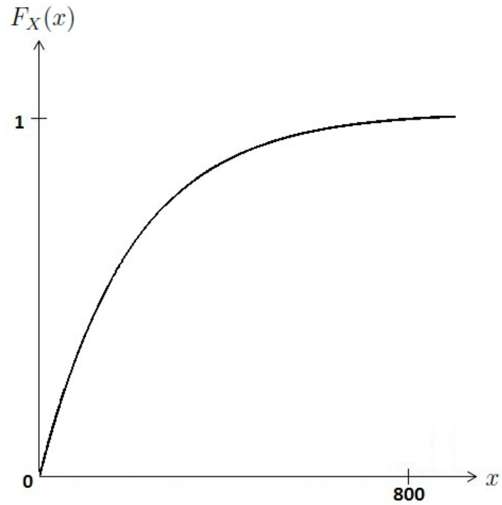

Figura 8.4: Função de distribuição acumulada de uma exponencial

(iv) Função de distribuição acumulada.

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \int_{-\infty}^{x} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

(v) Falta de memória. A distribuição exponencial possui a propriedade de falta de memória, ou seja, para quaisquer $s \geq 0$ e $t \geq 0$, tem-se que

$$P(X > s + t | X > s) = P(X > t).$$

Para verificar este fato, note que

$$P(X > s + t | X > s) = \frac{P(X > s + t, X > s)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)}.$$

Mas

$$P(X > s + t) = \int_{s+t}^{\infty} \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_{s+t}^{\infty} = e^{-\lambda(s+t)}.$$

Similarmente,

$$P(X > s) = e^{-\lambda s}.$$

Portanto,

$$P(X > s + t | X > s) = e^{-\lambda t} = P(X > t).$$

Exemplo 8.2.1: Observa-se que um tipo particular de *chip*, que tem duração de vida exponencial, é igualmente provável durar menos que 5000 horas ou mais que 5000 horas.

(a) Determine o tempo de duração médio de um *chip* deste tipo.

(b) Qual a probabilidade que o *chip* dure menos de 1000 horas ou mais de 10000 horas?

Solução: Seja X o tempo de duração deste *chip*. Para resolver o problema é preciso determinar seu parâmetro. Sabe-se que $P(X < 5000) = P(X > 5000)$, e como $P(X < 5000) + P(X > 5000) = 1$, tem-se que $P(X < 5000) = 0.5$. Portanto, $1 - e^{-\lambda(5000)} = 0.5$, ou seja, $\lambda = \frac{\log 2}{5000}$. Então, o tempo de duração médio deste tipo de *chip* é $\frac{5000}{\log 2}$ horas.

Para calcular a probabilidade pedida,

$$P((X < 1000) \cup (X > 10000)) = P(X < 1000) + P(X > 10000) = 1 - e^{-\frac{\log 2}{5}} + e^{-2\log 2}$$
$$= 1 - (2)^{-\frac{1}{5}} + (2)^{-2} = 1 - 0{,}8706 + 0{,}25 = 0{,}3794.$$

■

8.3 Normal de parâmetros μ e σ^2: $N(\mu, \sigma^2)$

(i) Densidade.

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}\ \mu \in \mathbb{R},\ \sigma \in \mathbb{R}^+.$$

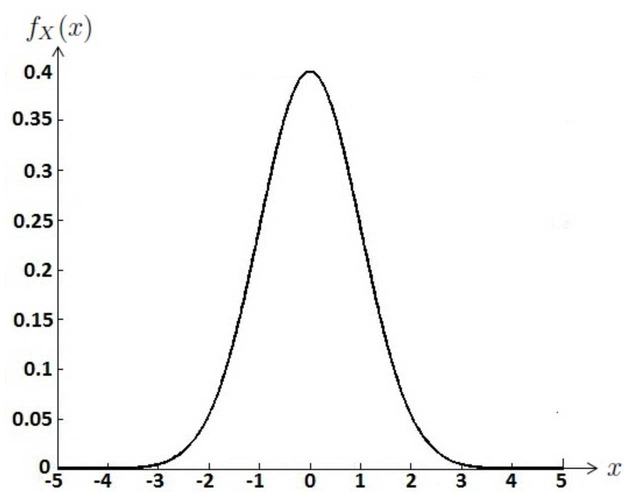

Figura 8.5: Densidade de probabilidade de uma normal

Para verificar que esta realmente é uma função densidade de probabilidade, realiza-se a seguinte substituição de variáveis $t = \frac{x-\mu}{\sigma}$, obtendo-se:

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} dt = I.$$

Para calcular I^2 utiliza-se o seguinte artifício.

$$I^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\frac{-t^2}{2}} dt \int_{-\infty}^{\infty} e^{\frac{-s^2}{2}} ds = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\frac{-(t^2+s^2)}{2}} dt ds.$$

Fazendo a mudança de variável $t = r\cos\theta$ e $s = r\,\text{sen}\,\theta$, tem-se:

$$I^2 = \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} re^{\frac{-r^2}{2}}drd\theta$$
$$= \frac{1}{2\pi}\int_0^{2\pi} -e^{\frac{-r^2}{2}}|_0^{\infty}d\theta$$
$$= \frac{1}{2\pi}\int_0^{2\pi} 1d\theta = 1.$$

Portanto, $I = 1$.

Segundo Fine (2006), historicamente esta distribuição foi chamada de "normal" porque era amplamente aplicada em fenômenos biológicos e sociais. Aplicações da distribuição normal incluem ruído térmico em resistores e em outros sistemas físicos que possuem um componente dissipativo; ruídos de baixa-frequência como os encontrados em amplificadores de baixa frequência; variabilidade em parâmetros de componentes manufaturados; comportamento de variáveis em organismos biológicos como, por exemplo, altura e peso.[1]

A densidade é simétrica em torno do parâmetro μ, e quanto menor o parâmetro σ mais concentrada é a densidade em torno de μ. Pode-se provar que os pontos $\mu - \sigma$ e $\mu + \sigma$ são os pontos de inflexão do gráfico de f_X. Será visto adiante que μ e σ^2 são a esperança e a variância da distribuição, respectivamente.

Se $\mu = 0$ e $\sigma^2 = 1$ esta densidade é chamada na literatura de *normal padrão* ou *normal reduzida*.

(ii) Esperança

$$E(X) = \int_{-\infty}^{\infty} x\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}dx.$$

Fazendo a mudança de variável $y = \frac{x-\mu}{\sigma}$, tem-se

$$E(X) = \int_{-\infty}^{\infty}\frac{\sigma y + \mu}{\sqrt{2\pi}}e^{\frac{-y^2}{2}}dy = \int_{-\infty}^{\infty}\frac{\sigma y}{\sqrt{2\pi}}e^{\frac{-y^2}{2}}dy + \int_{-\infty}^{\infty}\frac{\mu}{\sqrt{2\pi}}e^{\frac{-y^2}{2}}dy = 0 + \mu = \mu.$$

(iii) Variância.

Para o cálculo do segundo momento também é realizada a mudança de variável $y = \frac{x-\mu}{\sigma}$, logo

[1]Pode parecer estranho modelar quantidades que só assumem valores positivos por uma distribuição normal onde valores negativos aparecem. Nestes casos o que ocorre é que os parâmetros μ e σ^2 devem ser escolhidos de modo que a probabilidade da variável assumir um valor negativo seja aproximadamente nula o que torna a representação válida.

$$
\begin{aligned}
E(X^2) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma y+\mu)^2 e^{\frac{-z^2}{2}}dz \\
&= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2 e^{\frac{-z^2}{2}}dz + 2\mu\sigma\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z e^{\frac{-z^2}{2}} \\
&= +\mu^2\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\frac{-z^2}{2}}dz.
\end{aligned}
$$

A segunda parcela, pela resultado da esperança da normal padrão é igual a zero. A última parcela pelo resultado da integral da densidade da normal é igual a μ^2. Para calcular a primeira parcela usa-se integral por partes onde $u = z$ e $dv = ze^{\frac{-z^2}{2}}$, obtendo-se

$$
\begin{aligned}
E(X^2) &= \frac{\sigma^2}{\sqrt{2\pi}}(-ze^{\frac{-z^2}{2}}|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{\frac{-z^2}{2}}dz) + \mu^2 \\
&= \sigma^2 + \mu^2.
\end{aligned}
$$

O seguinte teorema afirma que transformações lineares de variáveis aleatórias com distribuição normal também são normalmente distribuídas.

Teorema 8.3.1: *Se $X \sim N(\mu, \sigma^2)$ e se $Y = aX + b$, onde $a > 0$ e $b \in I\!R$, então Y terá distribuição $N(a\mu + b, a^2\sigma^2)$.*

Prova: Note que

$$F_Y(y) = P(Y \leq y) = P(X \leq \frac{y-b}{a}) = F_X(\frac{y-b}{a}).$$

Derivando a expressão acima em relação a y,

$$f_Y(y) = \frac{1}{a}f_X(\frac{y-b}{a}) = \frac{1}{\sqrt{2\pi}a\sigma}e^{\frac{-(\frac{y-b}{a}-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}a\sigma}e^{\frac{(y-(b+a\mu))^2}{2a^2\sigma^2}},$$

ou seja, $Y \sim N(a\mu + b, a^2\sigma^2)$. ∎

Corolário 8.3.2: *Se $X \sim N(\mu, \sigma^2)$, então $Y = \frac{X-\mu}{\sigma}$ tem distribuição normal padrão.*

Pode-se provar que se $X_i \sim N(\mu_i, \sigma_i^2)$ são independentes, e $a_i \in I\!R$, para $i = 1, 2, 3, \ldots$, então $Y = c + \sum_{i=1}^{n} a_i X_i$ também tem distribuição normal com média $E(Y) = c + \sum_{i=1}^{n} a_i\mu_i$ e variância $V(Y) = \sum_{i=1}^{n}(a_i\sigma_i)^2$.

8.3.1 Tabulação da Distribuição Normal

Se $X \sim N(0,1)$, então

$$P(a < X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx.$$

Esta integral não pode ser resolvida analiticamente, contudo métodos numéricos de integração podem ser empregados para calcular integrais da forma acima e de fato valores de $P(X \leq z)$ existem em várias tabelas. A função de distribuição acumulada de uma normal padrão é usualmente denotada por Φ. Portanto,

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx.$$

Então, consultando valores de Φ em uma tabela, pode-se determinar que $P(a < X \leq b) = \Phi(b) - \Phi(a)$.

Utilizando o resultado do Corolário 8.3.2 e valores de Φ, pode-se obter para qualquer $X \sim N(\mu, \sigma^2)$, o valor de $P(a < X \leq b)$:

$$\begin{aligned} P(a < X \leq b) &= P(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}) \\ &= \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma}). \end{aligned}$$

Em especial o interesse pode ser calcular $P(\mu - k\sigma \leq X \leq \mu + k\sigma)$, usando o resultado acima tem-se que esta probabilidade é igual a $\Phi(k) - \Phi(-k)$.

Da simetria em torno de zero da normal padrão, segue-se que

$$\Phi(z) = P(X \leq z) = P(X \geq -z) = 1 - \Phi(-z)$$

para qualquer valor de z. Esta relação pode ser útil, pois frequentemente tabelas da distribuição normal padrão só possuem os valores positivos de s.

Exemplo 8.3.3: Suponha que X tenha uma distribuição $N(2, 0.16)$. Empregando uma tabela de distribuição normal calcule as seguintes probabilidades:

(a) $P(X \geq 2.3)$.

$$P(X \geq 2.3) = 1 - P(\leq 2.3) = 1 - \Phi(\frac{2.3-2}{0.4}) = 1 - \Phi(0.75) = 1 - 0.7734 = 0.2266.$$

(b) $P(1.8 \leq X \leq 2.1)$.

$$\begin{aligned} P(1.8 \leq X \leq 2.1) &= \Phi(\frac{2.1-2}{0.4}) - \Phi(\frac{1.8-2}{0.4}) = \Phi(0.25) - \Phi(-0.5) \\ &= 0.5987 - 0.3085 = 0.2902. \end{aligned}$$

■

Exemplo 8.3.4: Um equipamento com dois terminais com uma resistência equivalente de 1 Megohm opera em uma sala com temperatura de 300K. A voltagem térmica, V, que ele gera é observada na banda de 1.5GHz até 2.5GHz. Qual é a probabilidade de que a magnitude da voltagem exceda 8 milivolts? Assuma que $V \sim N(0, \sigma^2)$, onde $\sigma^2 = 4\kappa TRB$, κ é a constante de Boltzman que é igual a 1.38×10^{-23}, V é medido em volts, T é medido em graus Kelvin, R medido em ohms, e B medido em Hertz.
Solução: Das informações calcula-se que $\sigma^2 = 4(1.38 \times 10^{-23})(300)(10^6)(10^9) = 16.5 \times 10^{-6}$. Logo, $\sigma \approx 0.004$. Portanto,

$$\begin{aligned} P(|V| > 0.008) &= P(V > 0.008) + P(V < -0.008) = (1 - \Phi(\frac{0.008 - 0}{0.004})) + \Phi(\frac{-0.008 - 0}{0.004}) \\ &= 1 - \Phi(2) + \Phi(-2) = 2(1 - \Phi(2)) = 2(1 - 0.9772) = 0.456. \end{aligned}$$

■

8.4 Pareto de parâmetros α e k: Pareto(α, k)

(i) Densidade.

$$f_X(x) = \alpha k^\alpha x^{-(\alpha+1)}, k \in \mathbb{R}^+, x \geq k \ k \in \mathbb{R}^+.$$

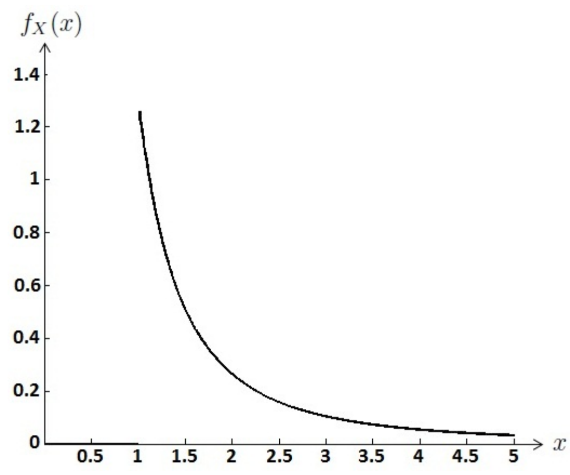

Figura 8.6: Densidade de probabilidade de uma distribuição de pareto

(ii) Esperança.

$$E(X) = \begin{cases} \alpha k(\alpha - 1)^{-1}, & \text{se } \alpha > 1, \\ \infty, & \text{se } \alpha \leq 1. \end{cases}$$

(iii) Variância.

$$V(X) = \begin{cases} \alpha k^2(\alpha-1)^{-2}(\alpha-2)^{-1}, & \alpha > 1, \\ \infty, & \alpha \leq 1. \end{cases}$$

(iv) Função de distribuição acumulada.

$$F_X(x) = \begin{cases} 0, & x < k, \\ 1-(\frac{k}{x})^\alpha, x \geq k. \end{cases}$$

Exemplo 8.4.1: Suponha que X tem uma distribuição de Pareto. Então,

(a) Se $k = 1$ e $\alpha = 0.25$, $P(1 < X < 2) = 0.16$.

(b) Se $k = 1$ e $\alpha = 1.3$, $P(1 < X < 2) = 0.88$.

■

A distribuição de Pareto pode ser utilizada para modelar distribuição de riquezas, atrasos em transmissão de pacotes e duração de sessões de internet, entre outros.

8.5 Weibull de parâmetros α e θ: $W(\alpha,\theta)$

(i) Densidade.

$$f_X(x) = \theta x^{(\theta-1)}\alpha e^{-\alpha x^\alpha},\ x \in \mathbb{R}^+,\ \alpha,\theta \in \mathbb{R}^+.$$

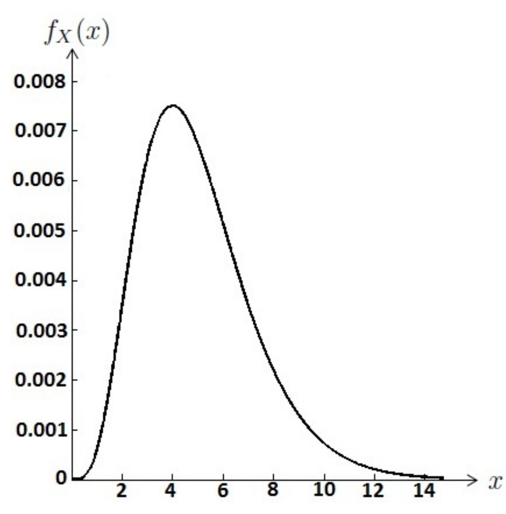

Figura 8.7: Densidade de probabilidade de uma weibull

(ii) Esperança.

$$E(X) = \alpha^{-1/\theta}\Gamma(\frac{1}{\theta} + 1).$$

(iii) Variância.

$$V(X) = \alpha^{-2/\theta}(\Gamma(\frac{2}{\theta} + 1) - (\Gamma(\frac{1}{\theta} + 1))^2).$$

8.6 Log-normal de parâmetros μ e σ^2: $LN(\mu, \sigma^2)$

(i) Densidade.

$$f_X(x) = (x\sigma\sqrt{2\pi})^{-1} exp\left(-\frac{1}{2}\frac{(\ln x - \mu)^2}{\sigma^2}\right), \; x \in \mathbb{R}^+, \; \sigma, \mu \in \mathbb{R}^+.$$

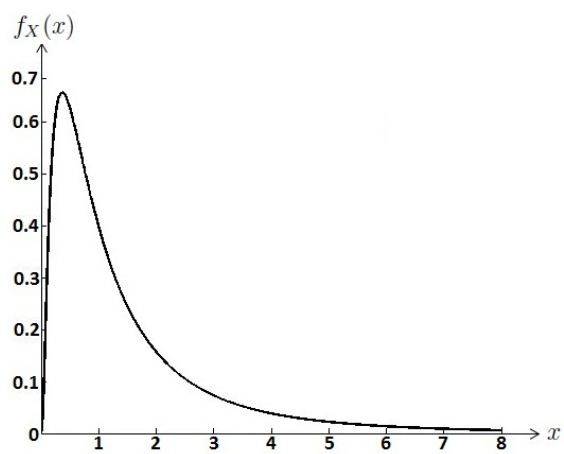

Figura 8.8: Densidade de probabilidade de uma log-normal

(ii) Esperança.

$$E(X) = \exp(\mu + \frac{1}{2}\sigma^2).$$

(iii) Variância.

$$V(X) = \exp(2\mu + \sigma^2(e^{\sigma^2} - 1).$$

8.7 Gama de parâmetros r e α: $G(r, \alpha)$

(i) Densidade.

$$f_X(x) = \frac{\alpha^r x^{r-1} e^{-\alpha x}}{\Gamma(r)}, \; x \in \mathbb{R}^+, r, \alpha \in \mathbb{R}^+,$$

onde

$$\Gamma(r) = \int_0^\infty x^{r-1} e^{-x}\, dx,\; r \in \mathbb{R}^+.$$

(ii) Esperança.

$$E(X) = \frac{r}{\alpha}.$$

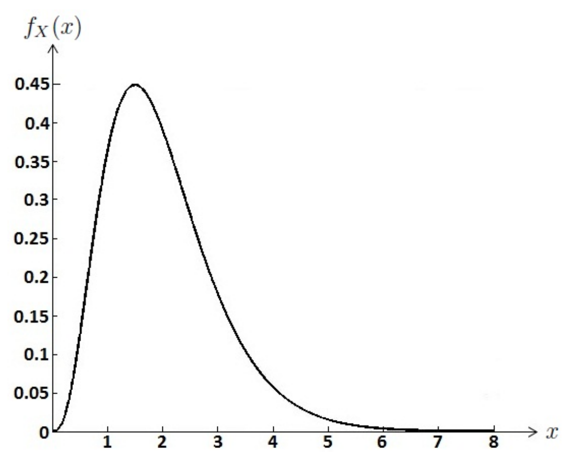

Figura 8.9: Densidade de probabilidade de uma distribuição gama

(iii) Variância.

$$V(X) = \frac{r}{\alpha^2}.$$

Com respeito à função Γ pode-se mostrar a seguinte relação de recorrência:

$$\Gamma(r) = (r-1)\Gamma(r-1).$$

Se $r = n$, onde n é um inteiro positivo,

$$\Gamma(n) = (n-1)!$$

Propriedades

(i) $r = 1 \Rightarrow G(1, \alpha) := Exp(\alpha)$.

(ii) Se $r = \frac{n}{2}$ e $\alpha = \frac{1}{2}$, tem-se uma distribuição χ^2 com n graus de liberdade.

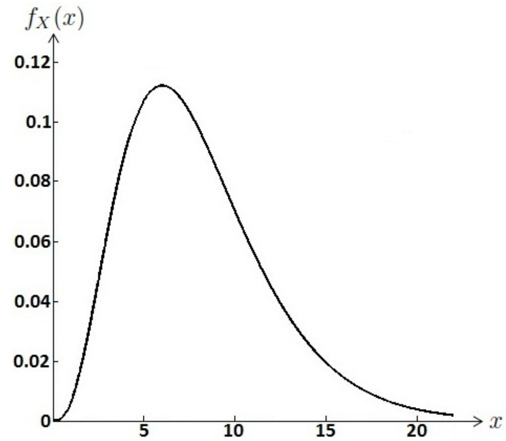

Figura 8.10: Densidade de probabilidade de uma distribuição qui-quadrado

8.8 Qui-quadrado de parâmetro n:, χ^2_n

(i) Densidade.

$$f_X(x) = \frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}, x \in \mathbb{R}^+,\ n \in Z^+.$$

(ii) Esperança.

$$E(X) = n.$$

(iii) Variância.

$$V(X) = 2n.$$

A distribuição χ^2_n tem tabeladas algumas probabilidades para determinados graus de liberdade.

Exemplos sobre o uso da tabela

(a) $P(\chi^2_9 \geq 16.9) = 5\%$.

(b) $P(\chi^2_6 \geq 1.24) = 97.5\%$.

(c) $P(\chi^2_{10} \geq x) = 98\% \Rightarrow x = 3.059$.

(d) $P(\chi^2_{30} \geq x) = 20\% \Rightarrow x = 36.250$.

(e) $P(\chi^2_{20} \leq x) = 0.05$. $P(\chi^2_{20} \leq x) = 1 - P(\chi^2_{20} \geq x) = 0.05 \Rightarrow P(\chi^2_{20} \geq x) = 0.95$ $\Rightarrow x = 10.851$.

(f) $P(\chi^2_{15} \leq x) = 10\%$. $P(\chi^2_{15} \geq x) = 90\% \Rightarrow x = 8.547$.

8.9 t-Student de parâmetro n: $t(n)$

(i) Densidade.

$$f_X(x) = \frac{\Gamma(\frac{n+1}{2})(1+\frac{x^2}{n})^{-(n+1)/2}}{\sqrt{2\pi}\Gamma(\frac{n}{2})}, \; x \in \mathbb{R}, \; n \in Z^+.$$

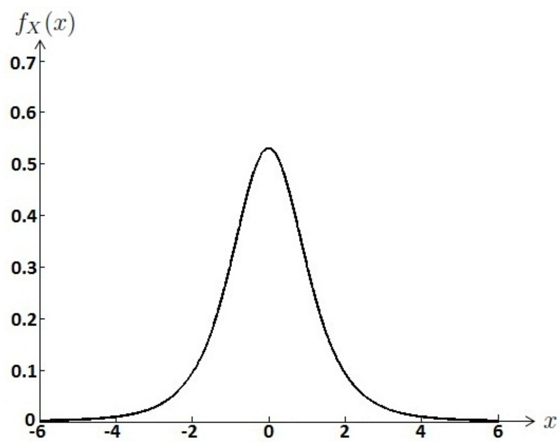

Figura 8.11: Densidade de probabilidade de uma distribuição t-student

(ii) Esperança.

$$E(X) = 0, \; n > 1.$$

$$\nexists E(X), n \leq 1,$$

(iii) Variância.

$$V(X) = \frac{n}{n-2}, n > 2.$$

Propriedades

(i) $\frac{N(0,1)}{\sqrt{\frac{\chi^2(n)}{n}}} \sim t(n)$.

(ii) $\lim_{n\to\infty} f_X(t) = N(0,1)$.

(iii) Se $n = 1$, tem-se que a distribuição t-Student é igual a uma distribuição Cauchy(0,1).

A distribuição t-Student é utilizada em inferência estatística. Por exemplo, pode-se utilizá-la para calcular um intervalo de confiança para a média de uma população quando a variância da população é desconhecida.

Exemplos de uso da tabela

(a) $P(-x < X_{10} < x) = 90\% \Rightarrow x = 1.812$.

(b) $P(-x < X_{29} < x) = 99\% \Rightarrow x = 2.756.$

(c) $P(X_{18} > x) = 2.5\% \Rightarrow x = 2.101.$

(d) $P(X_{25} < x) = 0.2\% \Rightarrow x = 2.166.$

8.10 F-Snedecor de parâmetros n e m: $F(n, m)$

(i) Densidade,

$$f_X(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \frac{(\frac{mx}{n})^{\frac{m}{2}-1}}{(1+\frac{mx}{n})^{\frac{m+n}{2}}} \frac{m}{n},\ x \in \mathbb{R}^+,\ n, m \in Z^+,$$

onde n e m são os parâmetros, $n, m \in Z^+$.

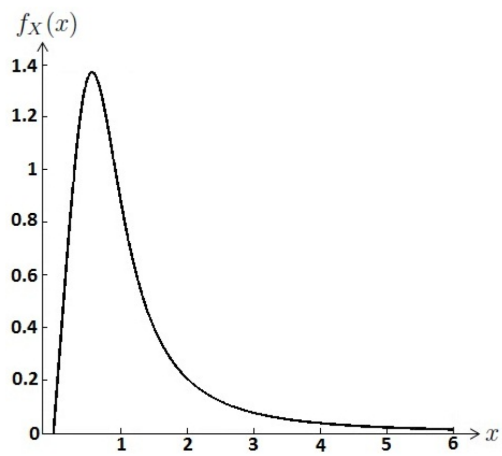

Figura 8.12: Densidade de probabilidade de uma distribuição f-snedecor

(ii) Esperança.

$$E(X) = \frac{n}{n-2}, n > 2.$$

(iii) Variância.

$$V(X) = \frac{2n^2(n+m-2)}{n(m-2)^2(m-4)}, m > 4.$$

Propriedades

(i)

$$F(n, m) = \frac{\chi^2(n)/n}{\chi^2(m)/m}.$$

(ii)

$$F(m, n) = \frac{\chi^2(m)/m}{\chi^2(n)/n} = \frac{1}{F(n, m)}.$$

(iii)

$$F(n,m,\alpha/2) = \frac{1}{F(m,n,1-\alpha/2)}.$$

Prova:

$$\begin{aligned} P(F(n,m) \leq a) &= \alpha/2 \Leftrightarrow \\ P(\frac{1}{F(m,n)} \leq a) &= \alpha/2 \Leftrightarrow \\ P(\frac{1}{a} \leq F(m,n)) &= \alpha/2 \Leftrightarrow \\ P(F(m,n) \geq \frac{1}{a}) &= \alpha/2 \Leftrightarrow \\ 1 - P(F(m,n) \leq \frac{1}{a}) &= \alpha/2 \Leftrightarrow \\ P(F(m,n) \leq \frac{1}{a}) &= 1-\alpha/2. \end{aligned}$$

Logo,

$$P(F(n,m) \leq a) = \alpha/2 \Leftrightarrow P(F(m,n) \leq \frac{1}{a}) = 1-\alpha/2.$$

Exemplos de uso da tabela

(a) $P(F(5,15) > x) = 5\% \Rightarrow x = 2.90.$

(b) $P(F(20,20) > x) = 5\% \Rightarrow x = 1.66.$

(c) $P(F(120,120) > x) = 5\% \Rightarrow x = 1.35.$

8.11 Aprendendo um pouco mais

8.11.1 Beta de parâmetros a e b: $Beta(a,b)$

(i) Densidade.

$$f_X(x) = \frac{x^{a-1}(1-x)^{b-1}}{\beta(a,b)}, x \in [0,1], a,b \in \mathbb{R}^+,$$

onde a e b são parâmetros, $a,b \in \mathbb{R}^+$ e $\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$.

(ii) Esperança.

$$E(X) = \frac{a}{a+b}.$$

(iii) Variância.

$$V(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Distribuições Beta são usadas exaustivamente em Estatística Bayesiana, pois são uma família de distribuições *a priori* conjugadas para distribuições binomiais e geométricas. A distribuição Beta pode ser utilizada para modelar eventos que têm restrição de estarem em um intervalo finito.

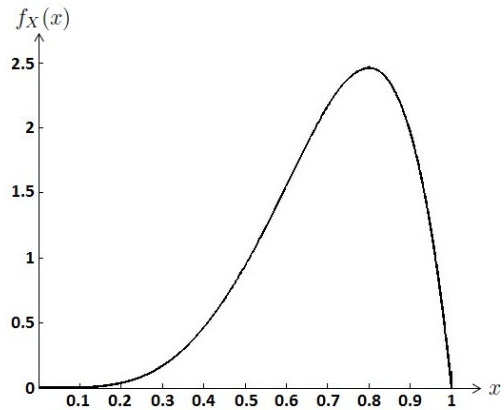

Figura 8.13: Densidade de probabilidade de uma distribuição beta

8.11.2 Cauchy de parâmetros x_0 e γ: $Cauchy(x_0, \gamma)$

(i) Densidade.

$$f_X(x) = \frac{1}{\pi} \cdot \frac{\gamma}{\gamma^2 + (x - x_0)^2}, x \in \mathbb{R},\ x_0 \in \mathbb{R}, \gamma \in \mathbb{R}^+.$$

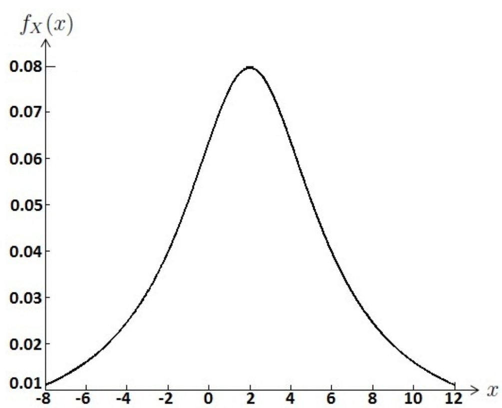

Figura 8.14: Densidade de probabilidade de uma distribuição cauchy

(iii) Esperança.

Se $X \sim Cauchy(x_0, \gamma)$, então X não é integrável, ou seja $E(X)$ não está definida, pois:

$$\int_{-\infty}^{0} \frac{x}{\pi} \cdot \frac{\gamma}{\gamma^2 + (x - x_0)^2} dx = -\infty,$$

e

$$\int_{0}^{\infty} \frac{x}{\pi} \cdot \frac{\gamma}{\gamma^2 + (x - x_0)^2} dx = \infty.$$

Prova-se (Meyer, 1983) que a razão entre duas variáveis aleatórias com distribuição normal independentes tem uma distribuição Cauchy com parâmetros $x_0 = 0$ e $\gamma = 1$.

8.11.3 A Distribuição Normal Bivariada

O vetor aleatório (X, Y) possui distribuição normal bivariada quando tem densidade dada por

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\{-\frac{1}{2(1-\rho^2)}[(\frac{x-\mu_1}{\sigma_1})^2 - 2\rho(\frac{x-\mu_1}{\sigma_1})(\frac{y-\mu_2}{\sigma_2}) + (\frac{y-\mu_2}{\sigma_2})^2]\},$$

onde $\sigma_1 > 0, \sigma_2 > 0, -1 < \rho < 1, \mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}$.

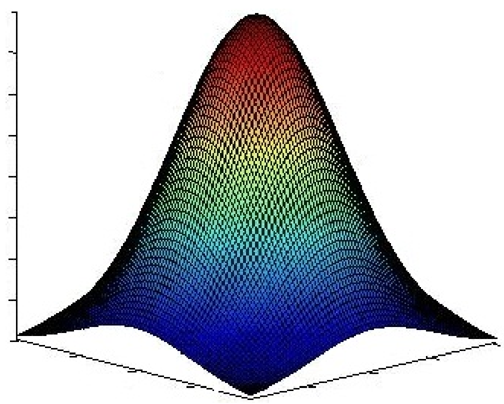

Figura 8.15: Densidade de probabilidade de uma distribuição normal bivariada

Se $\rho = 0$, esta densidade fatora e tem-se que X e Y são independentes. Se $\rho \neq 0$, esta densidade não fatora e X e Y não são independentes. Além disso, a distribuição normal bivariada satisfaz às seguintes propriedades:

(i) As distribuições marginais de X e de Y são $N(\mu_1, \sigma_1^2)$ e $N(\mu_2, \sigma_2^2)$, respectivamente.

(ii) O parâmetro ρ é igual ao coeficiente de correlação entre X e Y.

(iii) As distribuições condicionais de X dado $Y = y$ e de Y dado $X = x$ são, respectivamente,

$$N(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2), \sigma_1^2(1-\rho^2))$$

e

$$N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(y-\mu_1), \sigma_2^2(1-\rho^2)).$$

8.12 Exercícios

8.1 Como será a expressão da desigualdade de Tchebychev,

$$P(\mid X - \mu \mid \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

quando

(a) $X \sim B(n, p)$?

(b) $X \sim P(\lambda)$?

(c) $X \sim Exp(\alpha)$?

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

8.2 Compare o *limite superior* da probabilidade $P(\mid X - E(X) \mid \geq 2\sqrt{V(X)})$, obtido pela desigualdade de Tchebychev, com a probabilidade exata, em cada um dos seguintes casos:

(a) X é uniformemente distribuída sobre (-1,3).

(b) X tem distribuição $N(\mu, \sigma^2)$.

(c) X tem distribuição de Poisson com parâmetro 4.

(d) X tem distribuição exponencial com parâmetro α.

8.3 Uma variável aleatória contínua X é uniformemente distribuída no intervalo (α, β). Encontre a distribuição da variável $Y = -X$.

8.4 Se o número de sinais que, independentemente, chegam em um detector D no período de 10 segundos segue uma distribuição de Poisson com média 20, qual é a probabilidade de que chegue pelo menos um sinal em um período de 5 segundos?

8.5 Dois satélites idênticos são lançados simultaneamente ao espaço. A vida útil dos seus painéis solares pode ser modelada por uma distribuição exponencial de parâmetro 1 ano, isto é, as densidades são dadas por e^{-t}. O satélite A é monitorado durante os primeiros 2 anos e funciona perfeitamente; ele só tornará a ser verificado em algum momento após 5 anos de seu lançamento. (a) Qual a probabilidade de que ainda esteja funcionando nesse dia? O satélite B só é verificado em algum momento depois de completar 3 anos em órbita. (b) Qual a probabilidade de que ainda esteja funcionando nessa data? Compare essas probabilidades. (proposto por Thanius George R. Pinho)

8.6 Suponha que a duração de vida, T, de um dispositivo eletrônico, medida em horas, seja uma variável aleatória contínua com função densidade de probabilidade

$$f(t) = 0.01e^{-0.01t}, \ t > 0.$$

10 desses dispositivos são instalados independentemente em um sistema.

(a) Qual é a probabilidade de que um deles escolhido aleatoriamente dure menos de 50 horas?

(b) Qual é a probabilidade de que ao menos um dos 10 dispositivos dure menos de 50 horas?

8.7 A distribuição do tempo de acesso, T, a uma base de dados é normalmente distribuída com uma média de 5 msec e desvio padrão de 1 msec.

(a) Qual é a probabilidade de que este tempo ultrapasse 8 msec?

(b) Qual é a probabilidade de que este tempo seja inferior a 6 msec?

(c) Qual é o tempo t tal que com uma probabilidade de 0.95, o tempo de acesso seja menor que t?

8.8 O pessoal de uma firma de engenharia usa um terminal *on-line* para fazer seus cálculos. Sabe-se que o tempo de uso de um dado engenheiro segue uma distribuição exponencial com média 20 minutos. Qual é a probabilidade de que um engenheiro escolhido ao acaso

(a) passe menos de 30 minutos no terminal?

(b) ultrapasse a média da distribuição?

8.9 O tempo médio de CPU por sessão em sistemas *time-sharing* tem uma distribuição $N(4.4, 11.56)$. As sessões são classificadas como *trivial session* se tomam menos que 1 segundo de CPU, *editing session* se tomam entre 1 e 5 segundos de CPU, e *number-crunching session* em quaisquer outros casos.

(a) Calcule a probabilidade de cada tipo de sessão.

(b) Se 6 dessas sessões forem consideradas, qual é a probabilidade de que um igual número delas caia em cada uma das classificações acima?

8.10 Uma certa companhia telefônica taxa as chamadas da seguinte forma: \$0.20 para os primeiros 3 minutos; \$0.08 por minuto para qualquer tempo adicional. Suponha que o tempo de duração de uma chamada seja uma variável aleatória com distribuição exponencial de parâmetro 1. Defina a variável aleatória Y como sendo o custo por chamada.

(a) Estabeleça a distribuição de probabilidade de Y.

(b) Calcule o custo esperado.

8.11 Suponha que o número de mensagens no *buffer* em um sistema *on-line* tenha uma distribuição normal com média 100 e desvio padrão 10. Calcule a probabilidade de que o número de mensagens

(a) não exceda 120,

(b) esteja entre 80 e 120 e

(c) exceda 120.

8.12 O comprimento de uma determinada conexão, medida em centímetros, tem uma distribuição normal. A proporção de conexões com comprimento abaixo de 25 cm é 82%; a proporção de conexões com comprimento acima de 20 cm é 70%. Determine a proporção de peças que medem mais de 23 cm, sabendo-se que foram escolhidas entre aquelas que medem mais de 21 cm.

8.13 O tempo necessário para um estudante completar uma tarefa escolar tem distribuição normal com média 90 minutos e desvio padrão 15 minutos.

(a) Que proporção de estudantes termina a tarefa em 2 horas ou menos?

(b) Qual o tempo necessário para permitir que 90% dos estudantes terminem o teste?

(c) Em uma turma de 80 alunos, quantos se espera que terminem a tarefa em menos de 1 hora e 40 minutos, dentre os que a terminaram em mais de 1 hora e 10 minutos?

8.14 A duração de um certo tipo de pneu, em quilômetros rodados, é uma variável aleatória normal com duração média 60000km e desvio padrão de 10000km. Qual a probabilidade de que um pneu escolhido ao acaso dure

(a) mais de 75000km?

(b) entre 63000 e 70000km?

(c) O fabricante deseja fixar uma garantia de quilometragem, de tal forma que, se a duração do pneu for inferior à garantia, o pneu seja trocado. De quanto deve ser essa garantia para que a probabilidade de que o pneu seja trocado seja de 1% ?

8.15 Suponha que o tempo T entre chamadas em um dado sistema *on-line* tenha uma distribuição exponencial com um valor médio de 10 segundos.

(a) Encontre a variância de T.

(b) Qual é a probabilidade de que T não exceda 60 segundos?

(c) Exceda 90 segundos?

8.16 A duração da vida de um satélite é uma variável aleatória exponencialmente distribuída, com duração esperada de vida de 1.5 anos. Se três desses satélites forem lançados simultaneamente, qual será a probabilidade de que ao menos dois deles ainda venham a estar em órbita depois de 2 anos?

8.17 Quando um computador está operando, falhas ocorrem aleatoriamente. O tempo T até o aparecimento da primeira falha tem uma distribuição exponencial com parâmetro μ. Quando uma falha ocorre, é necessário corrigí-la dentro de um tempo t_0, depois do qual o computador começa a operar outra vez.

(a) Encontre a densidade e a função de distribuição do intervalo de tempo entre sucessivas falhas.

(b) Encontre a probabilidade de que este intervalo de tempo seja maior do que $2t_0$.

8.18 Suponha que o tempo T entre chamadas em um dado sistema *on-line* tenha uma distribuição exponencial com um valor médio de 10 segundos. Seja t um ponto arbitrário no tempo e X o tempo decorrido até a quinta chamada chegar (depois do tempo t). Encontre o valor esperado e a variância de X. Qual é a probabilidade de que T não exceda 60 segundos? Exceda 90 segundos?

8.19 A distribuição de Weibull,

$$F(x) = 1 - e^{-\alpha x^n},\ x > 0,$$

onde $\alpha > 0$ é uma constante e n é um inteiro positivo é frequentemente usada como a distribuição do tempo livre de falha de um equipamento.

(a) Encontre a função de densidade.

(b) Encontre sua média e variância.

8.20 Se X tem distribuição normal com média μ e variância σ^2, encontre b tal que

$$P(-b < (X - \mu)/\sigma < b) = 0.90.$$

8.21 Seja $X \sim N(\mu, \sigma^2)$ e tal que $P(X < 89) = 0.90$ e a $P(X < 94) = 0.95$. Encontre μ e σ^2.

8.22 Se $X \sim N(75, 25)$, encontre a probabilidade de que X seja maior do que 80 relativa à hipótese de que X seja maior do que 77.

8.23 Uma variável aleatória X tem distribuição normal com média μ e variância σ^2. Precisamos aproximar a distribuição normal por uma distribuição uniforme no intervalo (α, β) com os limites α e β escolhidos de tal forma que o valor médio e a variância de X sejam mantidos constantes. Determine α e β.

8.24 O tempo de vida de lâmpadas produzidas por uma certa fábrica segue uma distribuição exponencial com vida média de 200 horas.

(a) Qual a probabilidade de uma lâmpada escolhida ao acaso durar entre 200 e 300 horas?

(b) A fábrica deseja fixar uma garantia, de tal forma que se a duração da lâmpada for menor que a garantia, ela seja trocada. Qual deve ser a garantia do fabricante para repor apenas 5% da produção?

8.25 O comprimento das chamadas, T, em um *orelhão* de praia segue uma distribuição exponencial com média de 4 minutos. Os banhistas têm reclamado sobre a demora na fila, portanto a companhia telefônica resolve analisar o problema. Para tanto seleciona uma amostra de 50 usuários e calcula as seguintes probabilidades.

(a) da média amostral, $\overline{T}$, ser superior a 1 minuto;

(b) do mínimo amostral, K, ser superior a 1 minuto;

(c) do máximo amostral, M, ser inferior a 1 minuto.

Com base nos resultados obtidos, como você agiria?

8.26 Uma variável aleatória X tem uma distribuição de Simpson (obedece à lei de um triângulo isósceles) no intervalo $-a$ a a.

(a) Encontre a expressão da função densidade de probabilidade.

(b) Encontre sua média e variância.

(c) Encontre a probabilidade de que a variável aleatória assuma valores no intervalo $(-a/2, a)$.

8.27 Uma variável aleatória X tem distribuição exponencial com parâmetro μ.

(a) Encontre a função de distribuição acumulada e construa seu gráfico.

(b) Encontre a probabilidade de que a variável assuma um valor menor que sua média.

8.28 Uma variável aleatória X tem distribuição de Laplace,

$$f(x) = ae^{-\lambda|x|},$$

onde λ é um parâmetro positivo.

(a) Encontre o fator a.

(b) Construa o gráfico das funções de densidade e distribuição.

(c) Calcule sua média e variância.

8.29 A que transformação uma variável aleatória X uniformemente distribuída no intervalo $(0, 1)$ precisa ser submetida para que se obtenha, a partir de X, uma variável aleatória Y que tenha uma distribuição exponencial?

8.30 Suponha que X se distribui exponencialmente com parâmetro λ. Seja Y uma variável aleatória inteira definida em termos de X por $Y = m$ se $m \leq X < m + 1$, onde m é um número inteiro não negativo.

(a) Como se distribui Y?

(b) Calcule $E(Y)$.

8.31 A cada *ms*, o número de pacotes que chegam a um *switch* de uma rede segue uma distribuição de Poisson com parâmetro 2. No tempo dado, o *switch* só tem condição de gerenciar 15 pacotes. Portanto, se mais de 15 chegam, são excedentes e descartados.

(a) Qual é a probabilidade de haver descarte? Cálculos indicados.

(b) Qual é o número médio de pacotes que chegam ao *switch*?

(c) Qual é o tempo médio entre chegadas sucessivas de pacotes?

(d) Qual é a probabilidade de que o tempo entre chegadas sucessivas seja menor do que o tempo médio entre chegadas sucessivas?

(e) Suponha que 4 desses *switches* operem independentemente. Qual é a probabilidade de que haja descarte de pacotes em pelo menos 2?

Capítulo 9

Confiabilidade: Probabilidade aplicada

9.1 Introdução e Definições

Neste capítulo será vista uma aplicação dos conceitos relativos ao espaço das probabilidades apresentados anteriormente. Por meio dessa aplicação, denominada de **confiabilidade**, é possível realizar uma modelagem probabilística da forma com que um sistema falha, ou seja, a modelagem da *lei de falhas* do referido sistema. Assim, a confiabilidade é uma ferramenta utilizada para, entre outros, se prever quando um sistema poderá falhar pela primeira vez, conhecimento que é essencial para o planejamento de manutenções, especialmente quando o sistema analisado é crítico como um avião, por exemplo.

A Figura 1 ilustra o comportamento da incidência de falhas em um dado sistema, onde os intervalos de tempo em que o sistema permanece no estado *ON* e *OFF* indicam, respectivamente, a ausência ou não de falhas.

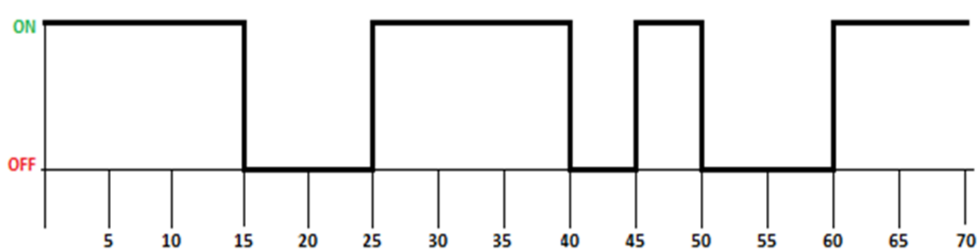

Figura 9.1: Comportamento da ocorrência de falhas

Confiabilidade é a probabilidade de um equipamento operar satisfatoriamente por um período de tempo dentro de condições previamente especificadas. Três aspectos devem ser levados em conta quando a confiabilidade de um sistema ou componente é analisada. Primeiro, uma definição *não ambígua* das possíveis falhas do sistema deve ser realizada. Em segundo lugar, a *unidade de tempo* deve ser identificada. Esta pode ser dada em horas, dias ou ciclos de operação do sistema. Para um sistema automotivo, por exemplo, a unidade de medida poderia ser ciclos de operação, representados pela ativação da ignição do motor. Por último, recomenda-se que o sistema seja *observado em condições de funcionamento natu-*

rais. A observação do sistema num ambiente manipulado pode gerar dados de confiabilidade viciados.

A forma com que um sistema falha é um fenômeno aleatório modelado por uma variável aleatória contínua T não negativa, que representa a duração de tempo até a sua primeira falha. Essa variável pode ser descrita por sua função densidade de probabilidade, $f_T(t)$, ou por sua função de distribuição acumulada, $F_T(t)$. Como T é contínua, então a probabilidade no ponto é zero. Ou seja, $P(T \geq t) = P(T > t)$.

Definição 9.1.1: Seja S um sistema observado em operação no período $[0, t]$. Então, a função confiabilidade de S, no instante t, $R(t)$, é dada por:

$$R(t) = P(T \geq t). \tag{9.1}$$

Com base em (9.1), tem-se que:

$$R(t) = \int_t^{+\infty} f_T(t)dt = 1 - \int_0^t f_T(t)dt. \tag{9.2}$$

A função densidade de probabilidade, $f_T(t)$, pode ser definida a partir da função confiabilidade. Partindo da Equação (9.2), tem-se que:

$$\frac{dR(t)}{dt} = -\frac{d \int_0^t f_T(x)dx}{dt}.$$

Assim,

$$f_T(t) = -\frac{dR(t)}{dt}. \tag{9.3}$$

É comum assumir que, para fins de cálculo da confiabilidade, o sistema opera satisfatoriamente no instante $t = 0$. Desse modo, $R(0) = 1$. Por outro lado, quando $t \to +\infty$, a confiabilidade é nula. A Figura 9.2 ilustra um exemplo de função confiabilidade, explicitando que $R(0) = 1$ e $\lim_{t \to +\infty} R(t) = 0$.

Proposição 9.1.2: *Seja T a variável que representa o tempo para falha de um sistema e $f_T(t)$ sua função densidade de probabilidade. Então, para valores $t = 0$ e $t \to +\infty$, a função confiabilidade $R(t)$ é dada, respectivamente, por:*

(a) $R(0) = 1$.

(b) $\lim_{t \to +\infty} R(t) = 0$.

Prova:

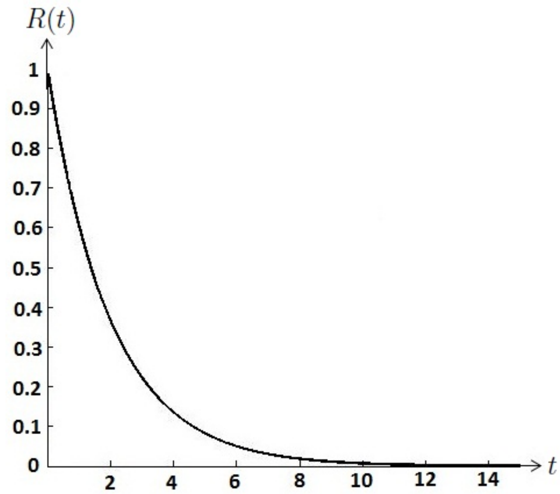

Figura 9.2: Exemplo da função confiabilidade

(a) Quando $t = 0$, tem-se que (9.2) é

$$R(0) = \int_0^{+\infty} f_T(t)dt. \tag{9.4}$$

Sabe-se que a T assume apenas valores não negativos, uma vez que $f_T(t)$ é uma função densidade de probabilidade, então:

$$\int_{-\infty}^{+\infty} f_T(t)dt = \int_0^{+\infty} f_T(t)dt = 1. \tag{9.5}$$

Assim, de acordo com (9.5),

$$R(0) = 1.$$

(b) Quando $t \to +\infty$, observa-se que (9.2) é dada por:

$$R(+\infty) = \lim_{t\to+\infty} \int_t^{+\infty} f_T(t)dt. \tag{9.6}$$

Considerando-se (9.6) tem-se que:

$$\lim_{t\to+\infty} \int_t^{+\infty} f_T(t)dt = \int_0^{+\infty} f_T(t)dt - \lim_{t\to+\infty} \int_0^{t} f_T(t)dt. \tag{9.7}$$

Da Equação (9.5), pode-se afirmar que:

$$\int_0^{+\infty} f_T(t)dt = 1$$

e

$$\lim_{t\to+\infty} \int_0^t f_T(t)dt = 1.$$

Dessa forma,

$$R(+\infty) = \lim_{t\to+\infty} \int_t^{+\infty} f_T(t)dt = 1 - 1 = 0.$$

∎

Outro conceito é o da *confiabilidade condicional*. Essa grandeza nada mais é do que a probabilidade que indica a chance de um sistema operar sem apresentar falhas no período de tempo $[t_0, t]$, dado que o referido sistema estava em funcionamento no momento $t = t_0$. Nesse caso, t_0 pode ser denominado como a idade do sistema.

Definição 9.1.3: A confiabilidade condicional é a seguinte probabilidade condicional:

$$\begin{aligned} R(t|t_0) &= P(T > t_0 + t | T > t_0) \\ &= \frac{P(T > t_0 + t)}{P(T > t_0)} \\ &= \frac{R(t_0 + t)}{R(t_0)},\ R(t_0) > 0. \end{aligned}$$

No contexto da análise da confiabilidade de sistemas, o *tempo médio para falha, de um sistema*, T_{med}, é dado pela esperança da variável aleatória T a qual especifica seu processo de falhas:

Definição 9.1.4:

$$T_{med} = E(T) = \int_0^{+\infty} t f_T(t)dt. \tag{9.8}$$

Em (9.8) tem-se que os limites da integral são 0 e $+\infty$ uma vez que T assume apenas valores não negativos.

Proposição 9.1.5: *Seja $R(t)$ a função confiabilidade de um determinado sistema. O Tempo Médio para Falha é dado por:*

$$T_{med} = \int_0^{+\infty} R(t)dt. \tag{9.9}$$

Prova: Substituindo (9.3) em (9.8) tem-se que:

$$T_{med} = \int_0^{+\infty} -\frac{dR(t)}{dt} t dt. \tag{9.10}$$

Empregando integração por partes em (9.10),

$$T_{med} = -tR(t)\Big|_0^{+\infty} + \int_0^{+\infty} R(t)dt. \tag{9.11}$$

Portanto,

$$\lim_{t\to+\infty} tR(t) = 0$$

e

$$0 \times R(0) = 0.$$

Ou seja,

$$T_{med} = 0 + \int_0^{+\infty} R(t)dt.$$

∎

Outra função utilizada na análise probabilística da confiabilidade de sistemas é a *taxa de falhas*. A taxa de falhas, $\lambda(t)$, é uma função que fornece a taxa instantânea de falhas de um sistema no momento t. A definição de $\lambda(t)$ baseia-se na seguinte probabilidade condicional:

$$P(t \le T \le t + \triangle t | T \ge t) = \frac{R(t) - R(t+\triangle t)}{R(t)}, \; R(t) > 0. \tag{9.12}$$

A Equação (9.12) define a chance de um sistema falhar em um período de tempo $\triangle t$, dado que não ocorreram falhas no intervalo de tempo $[0, t]$. Dessa forma, considerando (9.12), a probabilidade condicional de falhas por unidade de tempo é dada por

$$\frac{R(t) - R(t+\triangle t)}{R(t)\triangle t}. \tag{9.13}$$

A taxa de variação instantânea de falhas é dada pela probabilidade condicional de falhas por unidade de tempo, quando $\triangle t \to 0$. Assim,

$$\begin{aligned} \lambda(t) &= \lim_{\triangle t\to 0} \frac{-(R(t+\triangle t) - R(t))}{\triangle t} \times \frac{1}{R(t)} \\ &= \frac{-dR(t)}{dt} \times \frac{1}{R(t)} = \frac{f_T(t)}{R(t)}. \end{aligned} \tag{9.14}$$

Proposição 9.1.6: *Seja $\lambda(t)$ a taxa de falhas de um determinado sistema. Então, a função confiabilidade do sistema em questão é dada por*

$$R(t) = e^{-\int_0^t \lambda(t')dt'}. \tag{9.15}$$

Prova: De acordo com (9.14), tem-se que

$$\lambda(t) = \frac{-dR(t)}{dt} \times \frac{1}{R(t)}.$$

Dessa forma, pode-se definir

$$\lambda(t)dt = \frac{-dR(t)}{R(t)}. \tag{9.16}$$

Integrando a Equação (9.16), obtém-se:

$$\int_0^t \lambda(t')dt' = \int_{R(0)}^{R(t)} \frac{-dR(t')}{R(t')},$$

onde $R(0) = 1$. Portanto,

$$-\int_0^t \lambda(t')dt' = \ln R(t). \tag{9.17}$$

Assim,

$$R(t) = e^{-\int_0^t \lambda(t')dt'}. \tag{9.18}$$

∎

Exemplo 9.1.7: Seja a função densidade de probabilidade dada por:

$$f_T(t) = \begin{cases} 0.5e^{-0.5t}, & t > 0, \\ 0, & t \leq 0. \end{cases}$$

Nesse caso, a função confiabilidade é dada por:

$$R(t) = \int_0^{+\infty} 0.5e^{-0.5t}dt = e^{-0.5t}.$$

O valor do tempo médio para falha é obtido da seguinte forma:

$$T_{med} = \int_0^{+\infty} e^{-0.5t}dt = \frac{1}{0.5} = 2.$$

A taxa de falhas é calculada pela divisão da função densidade de probabilidade pela função confiabilidade. Assim, tem-se que:

$$\lambda(t) = \frac{f_T(t)}{R(t)} = \frac{0.5e^{-0.5t}}{e^{-0.5t}} = 0.5$$

A Figura 9.3 ilustra graficamente o comportamento de funções que caracterizam a confiabilidade do exemplo.

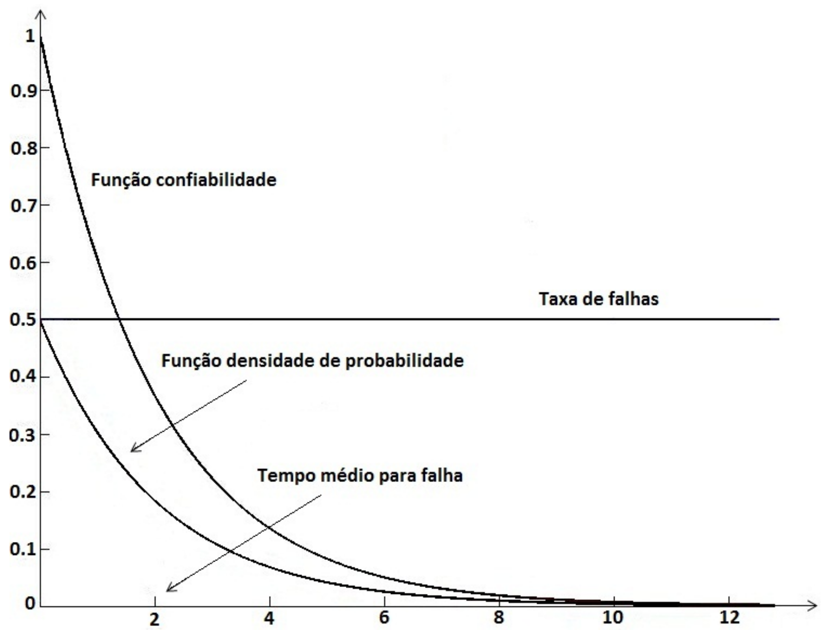

Figura 9.3: Gráficos de funções que caracterizam a confiabilidade do Exemplo 9.1.7.

■

9.2 Distribuições para Confiabilidade

Nesta seção serão descritas as funções densidade de probabilidade amplamente utilizadas na modelagem da lei de falhas de sistemas: Exponencial e Weibull.

A distribuição de falhas Exponencial resulta numa taxa de falhas constante, ou seja, a taxa de incidência de falhas permanece inalterada com o passar do tempo, indicando que o sistema em questão não sofre o desgaste do tempo. Essa distribuição é amplamente utilizada na modelagem de falhas de sistemas, uma vez que inúmeros sistemas possuem taxa de falhas constante. Por outro lado, com a distribuição Weibull é possível, por meio de variações de parâmetros dessa distribuição, tornar a taxa de falhas de um determinado sistema crescente ou decrescente, permitindo flexibilidade na modelagem.

9.2.1 Distribuição Exponencial para Falhas

Se a variável aleatória contínua possui distribuição Exponencial com parâmetro λ, então a função confiabilidade é dada por:

$$R(t) = \int_t^{+\infty} \lambda e^{-\lambda t} dt = e^{-\lambda t}.$$

Portanto a taxa de falhas é dada pela constante λ:

$$\lambda(t) = \frac{f_T(t)}{R(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$

Observa-se também que o tempo médio para falha, a esperança de T, é:

$$T_{med} = \int_0^{+\infty} R(t)dt = \int_0^{+\infty} e^{-\lambda t} dt = \frac{1}{\lambda}.$$

Sistemas modelados com a taxa de falhas constante são denominados sem memória[1]. Um sistema é denominado *instantâneo* (sem memória), caso as suas saídas em um instante de tempo t dependam somente de entradas geradas no mesmo instante t. Essa propriedade pode ser demonstrada empregando confiabilidade condicional:

$$R(t|T_0) = \frac{R(t+T_0)}{R(T_0)} = \frac{e^{-\alpha(t+T_0)}}{e^{-\alpha T_0}} = \frac{e^{-\alpha t} \cdot e^{-\alpha T_0}}{e^{-\alpha T_0}} = e^{-\alpha t} = R(t).$$

Em outras palavras, o tempo para falhas de um sistema sem memória depende apenas do tamanho do período de operação, não sendo afetado pela idade corrente do sistema (T_0), isto é, $R(t|T_0) = R(t)$.

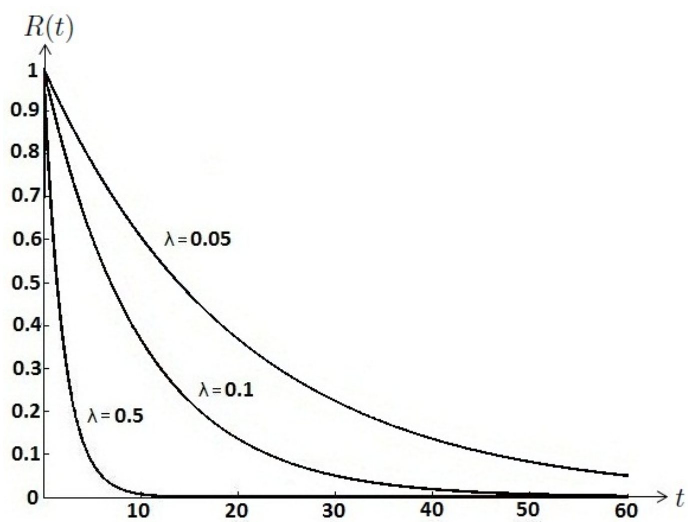

Figura 9.4: Gráficos de função confiabilidade para distribuição de falhas Exponencial.

[1]Falta de memória é um propriedade de sistemas com distribuição de falhas Exponencial.

9.2.2 Distribuição Weibull para Falhas

Como mencionado anteriormente, a distribuição Weibull é útil em confiabilidade, visto que a taxa de falhas para esse tipo de modelagem pode ser crescente ou decrescente. Nesse caso a taxa de falhas é dada por:

$$\lambda(t) = \frac{\alpha}{\theta}(\frac{t}{\theta})^{\alpha-1}, \tag{9.19}$$

$\alpha > 0$, $\theta > 0$, $t \geq 0$.

Utilizando (9.18) e considerando que $T \sim W(\alpha, \theta)$, tem-se que a função confiabilidade é:

$$R(t) = e^{[-\int_0^t \frac{\alpha}{\theta}(\frac{t}{\theta})^{\alpha-1}]} = e^{-(\frac{t}{\theta})^{\alpha}}. \tag{9.20}$$

Os parâmetros α e θ são, respectivamente, denominados como *parâmetro de forma* e *parâmetro de escala*. Para $\alpha \leq 1$, a função densidade de probabilidade é similar a uma exponencial, e para valores grande de α ($\alpha \geq 3$), $f_T(t)$ se assemelha a uma distribuição normal. Para $1 < \alpha < 3$, a função de densidade de probabilidade é assimétrica. Quando $\alpha = 1$, a densidade é idêntica a uma exponencial de parâmetro $\frac{1}{\theta}$. O parâmetro θ, por sua vez, afeta a média e a dispersão da distribuição.

Para a distribuição Weibull, o tempo médio para falha é dado por

$$T_{med} = \theta\Gamma(1 + \frac{1}{\alpha}), \tag{9.21}$$

onde $\Gamma(x) = \int_0^{+\infty} y^{x-1}e^{-y}dy$.

Exemplo 9.2.1: Seja $T \sim W(\alpha, \theta)$ e $\alpha = 4$ e $\theta = 1000$. Assim, a taxa de falhas é dada por:

$$\lambda(t) = \frac{4}{1000}(\frac{t}{1000})^3.$$

Nesse caso, a função confiabilidade é dada por:

$$R(t) = e^{-(\frac{t}{1000})^2}.$$

Para $t = 100$, a confiabilidade assume o seguinte valor

$$R(100) = e^{-(\frac{100}{1000})^2} = e^{-(\frac{1}{10})^2} = 0.99.$$

O tempo médio para falha possui o seguinte valor

$$E(T) = T_{med} = 1000\Gamma(1 + \frac{1}{4}) = 906.40.$$

■

9.3 Sistemas Complexos

Sistemas complexos ou compostos podem ser convenientemente representados como uma "caixa preta"que possui dois conjuntos de terminais acessíveis, os quais compõem a entrada e saída do sistema. Esses sistemas são formados por componentes que atuam juntos para realizar uma determinada função, a qual não seria possível de ser realizada com quaisquer partes individuais isoladas.

Neste capítulo assume-se que um componente pode apresentar apenas dois estados: *operacional* e *indisponível*. Nesse último estado, considera-se que o componente possui no mínimo uma falha que comprometa o seu funcionamento correto. Portanto, o estado de cada componente pode ser representado pela variável aleatória discreta K a qual assume apenas dois valores. Considere x_i o estado do i-ésimo componente de um dado sistema. Então, o valor de x_i é dado por

$$x_i = \begin{cases} 1, & \text{se o componente } i \text{ está operacional,} \\ 0, & \text{se o componente } i \text{ está indisponível.} \end{cases}$$

Seja o sistema complexo L possuindo n componentes. O vetor $\mathbf{x} = (\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n})$ representa o estado de todos os componentes de L. Assim como seus componentes, L também pode se apresentar em um dos estados de operação mencionados. Dessa forma, o estado de L, denotado por φ, é definido com base no vetor $\mathbf{x}$ por meio da seguinte função:

$$\varphi = \varphi(\mathbf{x}) = \varphi(\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}).$$

A primeira etapa do cálculo da confiabilidade de sistemas complexos envolve a obtenção desses valores de confiabilidade para cada um de seus componentes. Em seguida os valores obtidos são combinados de acordo com a configuração dos componentes, resultando no valor da confiabilidade referente ao sistema complexo como um todo. As seguintes configurações serão abordadas neste capítulo: **série** e **paralelo**. Observe que, para cada uma dessas configurações, uma função φ distinta será definida.

Para fins de cálculo, supõe-se que os componentes abordados nesta seção não são *reparáveis*, ou seja, quando um componente deixa de funcionar o modelo matemático proposto não considera ações de manutenção ou reparativas. Portanto, para sistemas não reparáveis, o tempo médio para falha é a duração de vida do sistema em questão.

9.3.1 Configuração em série

Neste tipo de arranjo em série, os componentes estão conectados logicamente um seguido do outro, como se observa na Figura 11.4. Então, para que um sistema complexo formado por componentes em série falhe, basta que apenas um de seus componentes individuais esteja indisponível.

Seja o sistema L formado por n componentes em configuração em série. A função φ que define o estado de L é dada por

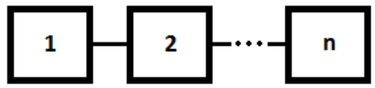

Figura 9.5: Representação da configuração em série

$$\varphi(\mathbf{x}) = \prod_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{x_i} = min(\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}). \tag{9.22}$$

A Equação (9.22) indica que o sistema complexo apresenta apenas um caminho crítico. Ou seja, caso L possua dois componentes, se apenas um deles falhar L também falhará.

Seja E_i o evento do i-ésimo componente não apresentar falhas durante um intervalo de tempo t. Tem-se que

$$P(E_i) = R_i(t),$$

onde $R_i(t)$ é a função confiabilidade do i-ésimo componente do sistema L. A função confiabilidade de L é dada por:

$$R(t) = P\left(\bigcap_{i=1}^{n} E_i\right). \tag{9.23}$$

Assumindo que os eventos E_i são independentes, (9.23) pode ser simplificada e representada da seguinte forma:

$$R(t) = P(E_1) \times P(E_2) \times \cdots \times P(E_n)$$

ou

$$R(t) = \prod_{i=1}^{n} R_i(t). \tag{9.24}$$

A taxa de falhas para a configuração com n componentes em série é baseada em (9.17). Demonstra-se que a taxa de falhas é dada unicamente com base na função confiabilidade do sistema em questão. Dessa forma, tem-se que

$$\lambda(t) = -\frac{d}{dt} lnR(t) = -\frac{d}{dt}\sum_{i=1}^{n} ln[R_{(i)}(t)]. \tag{9.25}$$

Utilizando (9.15), pode-se afirmar que a taxa de falhas para n componentes em série é

$$\lambda(t) = \sum_{i=1}^{n} \frac{f_{T(i)}(t)}{R_{(i)}(t)} = \sum_{i=1}^{n} \lambda_{(i)}(t), \tag{9.26}$$

onde $f_{T(i)}(t)$ e $\lambda_{(i)}(t)$ são a função densidade de probabilidade e taxa de falhas, respectivamente, referentes ao i-ésimo componente.

O tempo médio para falha para L pode ser definido utilizando a sua função confiabilidade $R(t)$. Então, o valor T_{med} é dado por

$$T_{med} = \int_0^{+\infty} R(t)dt. \tag{9.27}$$

Conforme (9.26), supondo que o sistema L possui n componentes apresentando lei de falhas Exponencial, a taxa de falhas de L é

$$\alpha_s = \sum_{i=1}^{n} \alpha_{(i)},$$

onde $\alpha_{(i)}$ é o parâmetro da distribuição exponencial do i-ésimo componente. Dessa forma, segundo (9.1), a função confiabilidade de L, nesse caso, pode ser representada por

$$R(t) = e^{-\alpha_s t}.$$

e o tempo médio para falha de L é

$$T_{med} = \frac{1}{\sum_{i=1}^{n} \lambda_{(i)}}, \tag{9.28}$$

onde $\lambda_{(i)}$ é a taxa de falhas do i-ésimo componente.

Exemplo 9.3.1: Seja L um sistema com três componentes, T_i a variável aleatória para o tempo de falha de cada elemento de L, onde $i = 1, 2$ e 3. Suponha que os componentes de L possuem processos de falhas exponencialmente distribuídos com parâmetros $\lambda_1 = 0.05$, $\lambda_2 = 0.1$ e $\lambda_3 = 0.2$, respectivamente. Então, o valor da função confiabilidade e tempo médio para falha do sistema L, considerando o instante $t = 5$, serão calculados.

$$R(5) = R_{\lambda_1}(5) \times R_{\lambda_2}(5) \times R_{\lambda_3}(5) = 0.17377394345045$$

$$T_{med} = \frac{1}{\lambda_1 + \lambda_2 + \lambda_3} = 2.857142857142857143$$

■

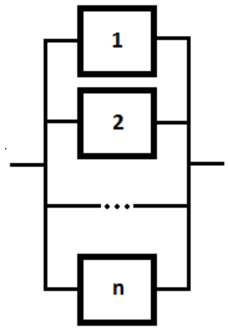

Figura 9.6: Representação da configuração em paralelo ou redundante

9.3.2 Configuração em paralelo

A configuração paralela, também denominada *redundante*, implica que o sistema com esta configuração falha apenas se todos os seus componentes se apresentarem em estado não operacional. A Figura 9.6 ilustra esta configuração.

Seja o sistema L formado por n componentes conectados em paralelo. A função φ que define o estado de L é dada por

$$\varphi(\mathbf{x}) = \mathbf{1} - \prod_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{1} - \mathbf{x_i}) = max(\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}). \tag{9.29}$$

A Figura 9.6 mostra que existem n caminhos críticos na configuração redundante ilustrada. Seja E_i o evento do i-ésimo componente não apresentar falhas durante um intervalo de tempo t. Portanto, a probabilidade do sistema L não falhar durante um período de tempo t é dada por

$$P(t \geq T) = P\left(\bigcup_{i=1}^{n} E_i\right).$$

Considerando independência dos processos de falhas entre os componentes, tem-se que

$$P(t \geq T) = 1 - \prod_{i=1}^{n} P(E_i^c), \tag{9.30}$$

onde E_i^c é o evento complementar de E_i.

Assim, a função confiabilidade de L, com base em (9.30), pode ser representada da seguinte forma:

$$R(t) = 1 - \prod_{i=1}^{n} [1 - R_{(i)}(t)], \tag{9.31}$$

onde $P(E_i^c) = 1 - R_{(i)}(t)$.

Seja o sistema Y com dois componentes conectados em configuração redundante, possuindo lei de falhas independentes com distribuição Exponencial com parâmetros α_1 e α_2. A função confiabilidade de tal sistema, sendo o mesmo observado por um período de tempo t, é denotada por

$$\begin{aligned} R(t) &= 1-(1-e^{-\alpha_1 t})\cdot(1-e^{-\alpha_2 t}) \\ &= e^{-\alpha_1 t}+e^{-\alpha_2 t}-e^{-(\alpha_1+\alpha_2)t}. \end{aligned} \tag{9.32}$$

Utilizando (9.9) e (9.32), o tempo médio para falha de Y é

$$\begin{aligned} T_{med}=\int_0^{+\infty} R(t)dt &= \int_0^{+\infty} e^{-\alpha_1 t}+\int_0^{+\infty} e^{-\alpha_2 t}-\int_0^{+\infty} e^{-(\alpha_1+\alpha_2)t} \\ &= \frac{1}{\alpha_1}+\frac{1}{\alpha_2}-\frac{1}{\alpha_1+\alpha_2}. \end{aligned} \tag{9.33}$$

Similarmente, o tempo médio para falha, considerando três componentes em configuração redundante, é

$$T_{med}=\frac{1}{\alpha_1}+\frac{1}{\alpha_2}+\frac{1}{\alpha_3}-\frac{1}{\alpha_1+\alpha_2}-\frac{1}{\alpha_1+\alpha_3}-\frac{1}{\alpha_2+\alpha_3}+\frac{1}{\alpha_1+\alpha_2+\alpha_3}.$$

Não existe uma fórmula fechada para o tempo médio considerando a configuração paralela. Caso se suponha que os n componentes possuem distribuição Exponencial com parâmetros idênticos, α, o tempo médio para falha é dado pela seguinte equação:

$$T_{med}=\frac{1}{\alpha}\sum_{i=1}^{n}\frac{1}{i}.$$

Exemplo 9.3.2: Seja L um sistema com três componentes em configuração paralela. Suponha que tais componentes possuem processo de falhas com distribuição Exponencial idênticas com parâmetro $\alpha = 0.04$. Então, o valor da função confiabilidade e tempo médio para falha, considerando o instante $t = 10$, serão calculados.

$$R(10)=1-(1-R_{\lambda=0.04}(10))\times(1-R_{\lambda=0.04}(10))\times(1-R_{\lambda=0.04}(10))=0.96416745766746$$

$$T_{med}=\frac{1}{0.04}+\frac{1}{0.04}+\frac{1}{0.04}=45.833333333333329$$

■

9.4 Exercícios

9.1 Considere um equipamente eletrônico com lei de falhas apresentando distribuição exponencial, tal que $T \sim E(0.01)$. Calcule:

(a) $R(0)$.

(b) $R(1.5)$.

(c) $R(10)$.

(d) $E(T)$.

9.2 Suponha que uma peça mecânica possui distribuição de falhas Weibull com parâmetros $\alpha = 5$ e $\theta = 400$. Calcule:

(a) $R(200)$.

(b) $\lambda(200)$.

(c) $R(400)$.

(d) $\lambda(400)$.

(e) $E(T)$

9.3 O tempo médio para falhas de um aparelho apresentando lei de falhas exponencial é dado por 1000 horas. Qual a confiabilidade desse equipamento após 10 horas de operação?

9.4 A duração da vida de um dispositivo eletrônico é exponencialmente distribiída. Sabe-se que a confiabilidade desse dispositivo para um período de 100 horas de operação é de 0.90. Quantas horas de operação devem ser levadas em conta para conseguir-se uma confiabilidade de 0.95?

9.5 Suponha um sistema com 2 componentes. O primeiro componente possui distribuição de falhas exponencial com taxa de falhas igual a 0.0001. O segundo possui distribuição de falhas Weibull com parâmetros $\alpha = 2$ e $\theta = 4000$. Calcule os seguintes valores para $t = 200$:

(a) $R(200)$, considerando que os componentes estão conectados em série.

(b) $R(200)$, considerando que os componentes estão conectados em paralelo.

9.6 Um sistema possui 2 componentes com distribuição de falhas independentes e conectados em série. O primeiro componente possui distribuição de falhas exponencial com taxa de falhas 0.0003. O segundo possui taxa de falhas dada por $\lambda(t) = \frac{t}{5\times 10^5}$. Encontre a confiabilidade do sistema após 100 horas de operação.

9.7 Suponha que a lei de falhas de um componente tenha a seguinte função densidade de probabilidade:

$$f(t) = \frac{(r+1)A^{r+1}}{(A+t)^{r+2}}, t > 0.$$

(a) Para quais valores de A e r, essa expressão é uma função densidade de probabilidade?

(b) Obtenha a expressão da função confiabilidade.

Capítulo 10

Teoremas Limite e Transformações de Variáveis Aleatórias

10.1 Introdução

Quando um problema do mundo físico deve ser resolvido num contexto de probabilidade, torna-se imprescindível conhecer a lei de probabilidade que rege o fenômeno que gerou o problema. Se o fenômeno pode ser enquadrado em alguma das situações abaixo descritas, é *fácil* encontrar a lei de probabilidade que o rege.

(i) Seja X uma variável aleatória com densidade $f_X(x)$, ou distribuição $F_X(x)$, Então, é fácil encontrar a lei de probabilidade de $Y = H(X)$.

(ii) Supondo que se tem o vetor aleatório (X, Y), com as leis de probabilidade de X e Y conhecidas. Também é fácil encontrar as distribuições de probabilidade de $X + Y$, $X - Y$, $X \times Y$ e X/Y, tanto considerando independência quanto dependência. Teoricamente, se se tem $(X_1, \ldots, X_n)$ *poderia* (mas não é!) ser fácil encontrar a distribuição de $(\ldots((X_1 + X_2) + X_3) + \cdots + X_n)$.

(iii) Se o vetor $(X_1, \ldots, X_n)$ é uniformemente distribuído numa região R^n de $\mathbb{R}^n$, encontrar a distribuiação de $(X_1, \ldots, X_n)$ envolve *apenas* um cálculo de volume (em $\mathbb{R}^n$), ficando as dificuldades técnicas por conta do cálculo abstrato e não de probabilidade.

(iv) Existem resultados, alguns aqui descritos na seção *Transformação de Variáveis Aleatórias*, que podem ser obtidos via o uso da função geratriz de momentos ou função característica.

Entretanto, se o problema a ser resolvido estrapola as situações anteriormente descritas, a solução é o uso de aproximações, as quais têm de ter um suporte teórico. Assim, os resultados vistos neste capítulo, portanto, têm como objetivo maior mostrar qual o suporte para realizar aproximações e quais as aproximações mais usadas no mundo real.

Um lembrete: o cálculo de limites na matemática tem como resultado o limite, quando este existe. No mundo probabilístico além da sintaxe, o cálculo efetivo de limite, tem uma semântica específica, isto é, *o experimento aleatório é ou foi realizado um número muito grande* *de vezes*, Mas, o que é grande? Esta resposta é um desafio na modelagem probabilística do mundo físico.

Neste capítulo será dada uma introdução a teoremas limite, teoremas relacionados com aproximações e resultados (ou teoremas) relativos a transformações de variáveis aleatórias. As demonstrações são omitidas uma vez que a maioria delas foge ao escopo do livro, entretanto, referências de onde encontrá-las serão fornecidas.

Por "teoremas limite" entende-se:

(i) Qual a distribuição da soma de n, onde esse n é grande, variáveis aleatórias independentes e igualmente distribuídas? E se as variáveis aleatórias não forem igualmente distribuídas?

(ii) Qual a relação entre a esperança matemática de uma variável aleatória e a média amostral, para uma amostra grande retirada independentemente dessa variável aleatória?

(iii) Foi visto no Capítulo 1 que a probabilidade do limite é o limite da probabilidade para sequências monotônicas de eventos. Será verdade que, dada uma sequência de variávies aleatórias, onde cada uma tenha esperança e o limite da sequência também tenha esperança, que o limite da esperança é a esperança do limite?

(iv) Qual a relação entre a probabilidade de um evento e sua frequência relativa ?

Os teoremas limite em probabilidade podem ser classificados como:

(i) *Lei de Grandes Números*, os quais analisam a estabilidade da média para um grande número de observações, e os

(ii) *Teoremas Centrais de Limite*, que tratam de distribuições assintóticas de probabilidade.

Estes teoremas fundamentam-se nos modos de convergência e na função característica, que são abordados na seção *Aprendendo um pouco mais*.

10.2 Lei de Grandes Números

(i) *Lei Fraca dos Grandes Números de Jakob Bernoulli (1713)*

Sejam $\mathcal{E}$ um experimento aleatório, A um evento associado a $\mathcal{E}$, n repetições independentes de $\mathcal{E}$, n_A o número de ocorrências de A nas repetições independentes de $\mathcal{E}$, $f_A = \frac{n_A}{n}$ e $P(A) = p$. Então

$$f_A \stackrel{\mathrm{P}}{\rightarrow} p,$$

isto é, f_A converge para p em probabilidade.

Prova: $E(f_A) = p$, $V(f_A) = \frac{p(1-p)}{n}$ porque $n_A \sim B(n,p)$. Usando a desigualdade de Tchebychev,

$$P(\mid f_A - p \mid < \epsilon) \geq 1 - \frac{p(1-p)}{n\epsilon^2} \Rightarrow$$

$$\lim_{n\to\infty} P(\mid f_A - p \mid < \epsilon) = 1. \tag{10.1}$$

∎

O resultado em (10.1) mostra qua a frequência de A quando n é grande pode ser uma aproximação para p. Adicionalmente, o modo de convergência envolvido é fraca, pois é convergência em probabilidade. Este teorema limite foi o primeiro da probabilidade e foi publicado em 1715.

Exemplo 10.2.1: O problema do administrador de uma rede é estimar a verdadeira proporção de usuários de uma dada aplicação usando uma amostra aleatória de tamanho n. Admitindo que a população de usuários é suficientemente grande para se usar um resultado assintótico, determine n para que o administrador possa garantir, com pelo menos 95% de probabilidade, que a distância entre a estimativa encontrada, p', e a probabilidade teórica 1/3, seja inferior a 3%.

Solução.

$$P(\mid p' - \frac{1}{3} \mid < 0.03) \geq 1 - \frac{(1/3)(2/3)}{n \times 0.03^2} = 0.95$$

Logo,

$$\frac{(1/3)(2/3)}{n \times 0.03^2} = 0.05 \Rightarrow n \geq 4938.$$

∎

(ii) *Lei Forte dos Grandes Números de Borel (1909)*

Sejam $X_1, X_2, \ldots, X_n, \ldots$ e X variáveis aleatórias independentes e identicamente distribuídas tais que $P(X_n = 1) = p$, $P(X_n = 0) = 1 - p$ e $S_n = X_1 + \cdots + X_n$. Então, $\frac{S_n}{n}$ converge em quase toda parte (qtp) para p, ou

$$\frac{S_n}{n} \stackrel{\text{qtp}}{\rightarrow} p.$$

Exemplo 10.2.2: Números Normais

Segundo Chung (1974), o teorema de Borel acima foi o primeiro caso de lei forte de grandes números e pode ser formulado em termos dos chamados *números normais*, como a seguir.

Seja $x \in [0, 1]$, escrito na expansão decimal (Kulisch e Miranker, 1981),

$$x = 0.d_1 d_2 \ldots.$$

Exceto por um conjunto contável de terminações decimais esta representação é única. Seja $0 \leq k \leq 9$ e v_k^n como sendo o número de dígitos entre os n primeiros dígitos de x que são iguais a k. Portanto, $\frac{v_k^n(x)}{n}$ é a *frequência relativa* do dígito k, nas n primeiras posições de x. Se existir $\varphi_k(x)$ tal que

$$\varphi_k(x) = \lim_{n \to \infty} \frac{v_k^n(x)}{n}$$

então $\varphi_k(x)$ pode ser chamada a frequência relativa de k em x.

x é chamado *simplemente normal* na base 10 se e só se, existe $\varphi_k(x)$ e

$$\varphi_k(x) = \frac{1}{10},\ 0 \leq k \leq 9.$$

Nestes termos, o teorema de Borel é: *exceto por um conjunto de Borel de medida nula, todo número em [0,1] é simplesmente normal.*

Resultados para a base 2 podem ser encontrados em James (1981).

Fazendo uma analogia com números de ponto-flutuante e gerador de números aleatórios em computadores, sabe-se que no intervalo [0,1] é onde existem mais números de ponto-flutuante (veja apêndice sobre o assunto) e que geradores de números aleatórios geram números aleatórios entre 0 e 1. Portanto, uma aplicação prática do teorema de Borel pode ser checar se o gerador de números aleatórios de uma determinada máquina é realmente aleatório, isto é, no limite, a frequência de cada dígito deveria ser $\frac{1}{2}$, uma vez que computadores trabalham internamente em binário, ■

10.3 Teoremas Centrais de Limite

(i) *Teorema Central do Limite de DeMoivre (1733)- Laplace (1812)*

Se $X \sim B(n,p)$ e $Z_n = \frac{X-np}{\sqrt{npq}}$, então $Z_n \xrightarrow{\text{D}} N(0,1)$.

Isto é,

$$\lim_{n\to\infty} P(Z_n \leq z) = \lim_{n\to\infty} F_n(z) = F_{N(0,1)}(z),$$

o que significa que Z_n converge em distribuição (D) para uma $N(0,1)$.

Este teorema mostra que uma $B(n,p)$ pode ser aproximada por uma $N(0,1)$. Ross (1993) sugere que a aproximação seja usada quando $npq \geq 10$.

Exemplo 10.3.1: Seja X o número de caras em 40 lançamentos independentes de uma moeda honesta. Calcular

(a) $P(X = 20)$.

(b) $P(X \geq 1)$.

Solução. Claramente X é uma binomial, porque o experimento é realizado independentemente um número finito de vezes e a probabilidade do evento de interesse, no caso aparecer cara, permanece constante durante a realização do experimento. Logo,

(a) $P(X = 20) = \binom{40}{20}(\frac{1}{2})^{20}(\frac{1}{2})^{40-20}$.

(b) $P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{40}{0}(\frac{1}{2})^{0}(\frac{1}{2})^{40}$.

O mesmo problema será resolvido aproximando a binomial pela $N(0,1)$. Para o cálculo de $P(X = 20)$, como a normal é contínua, probabilidade no ponto é zero, então usa-se o artifício de somar e subtrair 0.5 ao valor de interesse, no problema 20. Lembrando que $E(X) = 20$ e $V(X) = 10$.

(a) $P(X = 20) = P(19.5 \leq X \leq 20.5) = P(\frac{19.5-20}{\sqrt{10}} \leq \frac{X-20}{\sqrt{10}} \leq \frac{20.5-20}{\sqrt{10}}) = 0.1272.$

(b) $P(X \geq 1) = 1 - P(X < 1) = 1 - P(\frac{X-20}{\sqrt{10}} < \frac{1-20}{\sqrt{10}}) = 1 - P(Z_{40} < -6) = 1.$

■

(ii) *Teorema Central do Limite - TCL*

Sejam $X_1, X_2, \ldots, X_n, \ldots$, independentes tais que, para todo $i = 1, 2, \ldots,$, $E(X_i) = \mu_i$ e $V(X_i) = \sigma_i^2$. Seja $S_n = X_1 + \cdots + X_n$. Então,

$$Z_n = \frac{S_n - \sum^n \mu_i}{\sqrt{\sum^n \sigma_i^2}} \sim N(0,1).$$

Isto é,

$$\lim_{n\to\infty} P(Z_n \leq z) = \lim_{n\to\infty} F_n(z) = F_{N(0,1)}(z),$$

ou Z_n converge em distribuição para uma $N(0,1)$,

$$Z_n \xrightarrow{\mathrm{D}} N(0,1).$$

O teorema de DeMoivre-Laplace mostra que probabilidades envolvendo binomiais podem ser calculadas por meio de aproximação pela $N(0,1)$. Note que a convergência deste último resultado também é convergência em distribuição. O TCL fornece um método efetivo para se calcular probabilidades quando se tem somas de variáveis aleatórias independentes. Isto significa que se um fenômeno do mundo real puder ser modelado por uma soma (S_n) de n fatores independentes, mesmo não sendo possível encontrar uma fórmula para a distribuição de S_n, calcula-se qualquer probabilidade envolvendo S_n pela aproximação com a $N(0,1)$.

Exemplo 10.3.2: Sejam $X_i \sim U(0,1)$, para $i = 1, \ldots 10$, independentes. Calcular $P(\sum_1^{10} X_i < 7)$.

Solução. Sabe-se que $E(X_i) = \frac{1}{2}$ e $V(X_i) = \frac{1}{12}$. Seja $S_{10} = \sum_1^{10} X_i$. Logo, $E(S_{10}) = 5$ e $V(S_{10}) = \frac{10}{12}$. Portanto, $P(S_{10} < 7) = P(\frac{S_{10}-5}{\sqrt{\frac{10}{12}}} < \frac{7}{\sqrt{\frac{10}{12}}} = 0.9861.$ ■

10.4 Transformações de Variáveis Aleatórias

Nesta seção serão vistos resultados[1] relativos a aproximações de variáveis aleatórias por outras, ou de transformações de variáveis aleatórias.

(i) Se X se distribui como uma Hipergeométrica de parâmetros N, n e r , se $p = \frac{r}{N}$, para N grande,

$$P(X = k) \simeq B(n, p).$$

(ii) Sejam $X_1, X_2, \dots, X_n$ independentes tais que $X_i \sim N(\mu_i, \sigma_i^2)$, para $i = 1, \cdots, n$. Seja $X = X_1 + X_2 + \cdots + X_n$. Então,

$$X \sim N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2).$$

(iii) Sejam $X_1, X_2, \dots, X_n$ variáveis aleatórias independentes tais que $X_i \sim Poisson(\lambda_i)$, para $i = 1, \cdots, n$. Seja $X = X_1 + X_2 + \cdots + X_n$. Então,

$$X \sim Poisson(\sum_{i=1}^{n} \lambda_i).$$

(iv) Se $X_k \sim B(n_k, p)$, para $k = 1, \cdots, n$ e $X = X_1 + X_2 + \cdots + X_n$. Então,

$$X \sim B(\sum_{i=1}^{n} n_k, p).$$

(v) Sejam $X_1, X_2, \dots, X_n$ independentes tais que $X_i \sim N(\mu_i, \sigma_i^2)$, para $i = 1, \cdots, n$ e seja $\overline{X} = \frac{X_1+X_2+\cdots+X_n}{n}$. Então,

$$\overline{X} \sim N(\frac{\sum_{i=1}^{n} \mu_i}{n}, \frac{\sum_{i=1}^{n} \sigma_i^2}{n^2}).$$

(vi) Sejam $X_1, X_2, \dots, X_n$ variáveis aleatórias independentes tais que $X_i \sim Exp(\alpha)$, para $i = 1, \cdots, n$ e seja $G = X_1 + X_2 + \cdots + X_n$. Então,

$$G \sim Gama(\alpha, n).$$

(vii)

$$\chi_n^2 = \sum_{i=1}^{n} X_i^2, X_i \sim N(0, 1), \text{ independentes.}$$

(viii) *Sejam $X_1, \cdots, X_n$ variáveis aleatórias independentes tais que $X_i \sim \chi^2_{n_i}$ para $i = 1, \cdots, k$ e $X = \sum_{i=1}^{k} X_i$. Então, $X \sim \chi^2_n$, onde $n = n_1 + \cdots + n_k$.*

[1]Provas via função geratriz de momentos ou função caracterítica (ver a última seção).

10.5 Aprendendo um pouco mais

Nesta seção serão vistas as definições dos modos de convergência e da função característica e duas leis de grandes números que ilustram convergência em probabilidade e em quase toda parte.

10.5.1 Modos de Convergência

(i) Convergência *em quase toda parte*, ou *quase certa* ou *com probabilidade 1.*

Sejam $X_1, X_2, \ldots, X_n, \ldots$ e X variáveis aleatórias em $(\Omega, \mathcal{A}, P)$. X_n converge para X em quase toda parte,

$$X_n \stackrel{\text{qtp}}{\to} X$$

se

$$P(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$$

Este resultado significa que o conjunto dos ω onde $X_n \not\to X$ tem probabilidade zero. Este é o significado de convergñcia em quase toda parte: convergência fora de um conjunto de probabilidade (medida) nula. Adicionalmente, convergência em quase toda parte é convergência pontual fora de um conjunto de medida nula.

(ii) Convergência *em probabilidade*

Sejam $X_1, X_2, \ldots, X_n, \ldots$ e X variáveis aleatórias em $(\Omega, \mathcal{A}, P)$. X_n converge para X em probabilidade,

$$X_n \stackrel{\text{P}}{\to} X$$

se, $\forall \epsilon > 0$,

$$\lim_{n\to\infty} P(\{\omega :\mid X_n(\omega) - X(\omega) \mid > \epsilon\}) = 0.$$

A semântica da convergência em probabilidade é: quando n é grande, isto é, quando o experimento é realizado um número muito grande de vezes os valores de $X_n(\omega)$ e $X(\omega)$ são probabilisticamente os mesmos.

(iii) Convergência *em distribuição*

Sejam as variáveis aleatórias $X_1, X_2, \ldots, X_n, \ldots$ e X não necessariamente no mesmo espaço de probabilidade $(\Omega, \mathcal{A}, P)$. X_n converge para X em distribuição,

$$X_n \stackrel{\text{D}}{\to} X$$

se, para todo x ponto de continuidade de $F_X(\cdot)$,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x).$$

Além desses, existem outros modos de convergência (por exemplo, convergência em média de ordem p, $0 < p < +\infty$), entretanto, os aqui citados são suficientes para o entendimento de qual modo de convergência foi usado neste capítulo.

Convergência em quase toda parte é denominada de convergência *forte* e em probabilidade de *fraca*. O mais fraco modo de convergência é em distribuição. Pode-se mostrar que:

$$\text{qtp} \Rightarrow \text{P} \Rightarrow \text{D}.$$

10.5.2 Função Característica

Seja X uma variável aleatória. A função característica de X, φ, é

$$\varphi : \mathbb{R} \to \mathbb{C}$$

$$\varphi_X(t) = E(e^{itX}).$$

Portanto, φ_X toma valores complexos. Na verdade, φ_X é a esperança matemática da função da variável aleatória X, e^{itX}.

A função caractarística tem várias propriedades, entre outras, ele determina a distribuição acumulada de X, F_X, e é determinada por ela. Na verdade, esta propriedade é um teorema denominado *teorema da unicidade* e sua tese é que se $\varphi_{X_1}(t) = \varphi_{X_2}(t)$, para todo t, então $F_{X_1}(x) = F_{X_2}(x)$, para todo x.

A versão real da função caractarística é *função geratriz de momentos* ou *função de momentos*, M_X,

$$M_X : \mathbb{R} \to \mathbb{R}$$

$$M_X(t) = E(e^{tX}).$$

O *teorema da unicidade* também é válido se $\varphi_X(t)$ é substituída por $M_X(t)$.

10.5.3 Lei Fraca dos Grandes Números de Khintchine (1929)

Sejam $X_1, X_2, \ldots, X_n, \ldots$ e X variáveis aleatórias independentes e identicamente distribuídas em $(\Omega, \mathcal{A}, P)$ tais que $E(X_i) = \mu < \infty$, $\forall i$ e $S_1, S_2, \ldots, S_n = X_1 + \cdots + X_n, \ldots$ as somas parciais. Então,

$$\frac{S_n}{n} \stackrel{\text{P}}{\to} \mu.$$

10.5.4 Lei Forte dos Grandes Números de Kolmogorov (1933)

Sejam $X_1, X_2, \ldots, X_n, \ldots$, e X variáveis aleatórias independentes e identicamente distribuídas tais que $E(X_i) = \mu < \infty$, $\forall i$. Seja $S_n = X_1 + \cdots + X_n$. Então,

$$\frac{S_n}{n} \stackrel{\text{qtp}}{\to} \mu.$$

10.6 Exercícios

10.1 Sejam $X_k \sim Poisson(1)$, $k = 1, \cdots, 10$, independentes. Calcular $P(\sum_{k=1}^{10} X_k \geq 15)$.

10.2 Determinados programas que chegam a um sistema de computação requerem um tempo de CPU que pode ser modelado por uma distribuição exponencial com parâmetro $1/140$ milissegundos. Por razões operacionais, a disciplina da CPU é tal que, se um programa não for processado dentro de 100 milissegundos, sua execução é interrompida e o programa volta para o final da fila de acesso à CPU.

(a) Encontre a probabilidade de que um desses programas volte para o final da fila.

(b) Suponha que um analista está interessado no tempo total, T, requerido à CPU por 40 desses programas. Qual é a probabilidade de que T ultrapasse 6000 milissegundos?

10.3 Adicione 100 números reais, cada um dos quais é arredondado para o inteiro mais próximo. Assuma que cada erro de arredondamento é uma variável aleatória uniformemente distribuída entre -0.5 e 0.5, e que estes erros sejam independentes.

(a) Encontre, aproximadamente, a probabilidade de que o erro na soma esteja entre -3 e 3.

(b) Encontre a quantidade x tal que, com aproximadamente 99% de probabilidade, em valor absoluto o erro na soma seja menor do que x.

10.4 Sejam $l_1, l_2, \cdots, l_n, \cdots$ lâmpadas quaisquer instaladas de tal forma que l_2 começa a funcionar quando l_1 deixa de funcionar, l_3 começa a funcionar quando l_2 deixa de funcionar, e assim por diante. Assume-se que cada lâmpada tem um tempo médio de vida de 2 meses com um desvio padrão de 0.25 mês.

(a) Encontre, aproximadamente, a probabilidade de que 40 lâmpadas durem pelo menos 7 anos.

(b) Quantas lâmpadas, n, devem ser instaladas de formas que, com uma probabilidade de 0.95 as n lâmpadas durem pelo menos 5 anos?

10.5 Um avião de turismo de 4 lugares pode levar uma carga útil de $350kg$. Carga útil = peso dos 4 passageiros + o peso das respectivas bagagens. Sabe-se que o peso de qualquer dos passageiros segue uma distribuição normal com média $70kg$ e variância $400kg^2$, e de qualquer respectiva bagagem, também normal com média $12kg$ e variância $25kg^2$.

(a) Qual é a probabilidade de haver sobrecarga se o piloto, irresponsavelmente, não pesar os 4 passageiros e as respectivas bagagens?

(b) Qual é a probabilidade de que o piloto tenha de descartar pelo menos $50kg$ de combustível para evitar sobrecarga?

10.6 48 medições são registradas com várias casas decimais. Cada um desses 48 números é arredondado para o inteiro mais próximo. A soma dos 48 números originais é aproximada pela soma destes inteiros. Assumindo que os erros de arredondamento são estocásticamente independentes e têm distribuição uniforme no intervalo $(-\frac{1}{2}, +\frac{1}{2})$, compute um valor aproximado para a probabilidade de que o erro seja menor que 2 unidades.

10.7 A espessura de trilhas fotorresistivas (pastilhas) na fabricação de semicondutores tem média e variância, respectivamente, 10 micrômetros e 2 micrômetros. Considere normalidade. Também considere que as espessuras de diferentes pastilhas sejam independentes.

(a) Determine a probabilidade de a espessura média de 10 pastilhas ser maior do que 11 ou menor do que 9 micrômetros.

(b) Qual deve ser o tamanho da amostra, n, tal que, com probabilidade 0.01, a espessura média seja maior do que 10 micrômetros?

10.8 Em um extenso programa de computador, o programador decide manter J algarismos significativos após o ponto decimal, arredondando o resultado de qualquer operação aritmética para este número de algarismos. Admita que haja um total de 10^6 operações elementares envolvendo arredondamentos, que os erros de arredondamento sejam independentes e uniformemente distribuídos no intervalo $[-0.5 \times 10^{-J}, 0.5 \times 10^{-J}]$. Calcule a probabilidade de que o erro de arredondamento seja menor do que 500×10^{-J}.

10.9 O peso de um equipamento eletrônico se distribui como uma normal com média 10g e desvio padrão 0.5g. Este equipamento é embalado em caixas contendo 120 unidades. O peso da caixa é 150g. Qual é a probabilidade de que uma caixa cheia pese mais de 1360g?

10.10 Sejam 27 voltagens independentes $V_1, \cdots, V_{27}$ recebidas por um somador $V = \sum_{i=1}^{27} V_i$. Suponha que $V_i \sim U(0, 10)$, $\forall i = 1, \cdots, 27$.

(a) Qual é a probabilidade de que a voltagem na entrada do somador exceda 148.5 volts?

(b) Sabe-se que a capacidade máxima desse somador é 200 volts. Calcule o número máximo, n, de voltagens que o mesmo pode receber de modo que sua capacidade não seja ultrapassada em pelo menos metade das vezes.

10.11 Supondo que a distribuição de probabilidade da carga útil de cada passageiro, seu peso mais sua bagagem, é uma Normal de média $82kg$ e variância $49kg^2$ e que um avião de 16 lugares pode levar uma carga útil total de $1360kg$, calcule a probabilidade de que o piloto tenha que tirar pelo menos $25kg$ de gasolina para evitar sobrecarga.

10.12 Um analista tem de entrevistar 20 programadores. Ele sabe que o tempo gasto em uma entrevista, T, segue uma lei normal com média 10 minutos e desvio padrão 3 minutos. Ele começa as entrvistas às 9:00.

(a) Qual é a probabilidade de que ele termine as entrevistas antes das 12:30 horas?

(b) Qual é a probabilidade de que ele termine as entrevistas antes das 12:30 horas, se parou durante 15 minutos para um cafezinho?

(c) Ele começa as entrevistas às 9:00 horas e decide parar para um cafezinho no tempo t, onde t é tal que, com 99% de certeza ele terá entrevistado 50% dos programadores. Quanto é t, isto é, quando ele parará?

10.13 Um produto pesa em média 8g com um desvio padrão de 5g. É embalado em caixa de 144 unidades que pesam em média 200g com desvio padrão 10g. Calcular a probabilidade de que uma caixa cheia pese mais do que 1400g, admitindo distribuições normais dos pesos.

10.14 A quantidade de tempo que um consumidor gasta esperando no balcão de *check-in* de um aeroporto é uma variável aleatória com média de 8.2 minutos e desvio-padrão de 1.5 minuto. Suponha que uma amostra aleatória de 49 consumidores seja observada. Encontre a probabilidade de que o tempo médio de espera na fila para esses consumidores esteja entre 5 e 10 minutos.

10.15 A vida efetiva de um componente usado em um motor de uma turbina de um avião a jato é uma variável aleatória com média 5000 horas e desvio-padrão 40 horas. Suponha normalidade e independência onde necessário. O fabricante do motor introduz uma melhoria no processo de fabricação desse componente que aumente a vida média para 5050 horas e diminui o desvio padrão para 30 horas. Uma amostra aleatória de 16 componentes é selecionada do processo antigo e outra de 25 elementos é selecionada do processo novo. Qual é a probabilidade de que a diferença entre as duas médias amostrais seja no mínimo de 25 horas?

10.16 Um sinal, S, recebido em um detector, D, medido em microvolts, em determinado instante t, pode ser modelado por uma $N(200, 256)$. Realizando cálculos com duas decimais e arredondamento simétrico, responda às questões a seguir.

(a) Qual é a probabilidade de que a voltagem do sinal exceda 220 microvolts?

(b) Se 10 desses sinais chegam independentemente, qual é a probabilidade de que ao menos um deles exceda 220 microvolts?

(c) Qual é a probabilidade de que o primeiro sinal que exceda 220 microvolts seja o décimo? Os sinais são independentes.

(d) Suponha que 30 sinais chegam alternadamente e independentemente ao detector. Qual a probabilidade de que a voltagem total, V, entre a chegada do primeiro e o recebimento do último sinal exceda 6000 microvolts?

(e) Um técnico constata que o número de sinais que chegam em D no período de 20 segundos segue um processo de Poisson com taxa 8. Qual é a probabilidade que cheguem 5 sinais em 30 segundos?

(f) A confiabilidade, R, do detector D é mensurada pela sua probabilidade de detectar sinais que excedam 220 microvolts: se $R \geq 0.5$, D é confiável; caso contrário, não. O cálculo proposto para a confiabilidade de D foi: D é confiável se é necessário se enviar, independentemente 5 sinais a fim de se detectar 2 que excedam 220 microvolts. D é confiável? Justifique sua resposta.

10.17 A capacidade máxima de um elevador é 2000kg. Supondo que o peso de um passageiro tem distribuição N(70kg, 100kg^2), use a **média amostral** para calcular:

(a) a probabilidade de 30 passageiros pesarem além do limite;

(b) o número máximo **n** de passageiros, de modo que a capacidade não seja ultrapassada em pelo menos metade das vezes.

10.18 Amostras independentes de tamanhos 10 e 15 são tiradas de uma variável aleatória normalmente distribuída, com esperança 20 e variância 3. Qual é a probabilidade de que as médias difiram, em valor absoluto, em mais de 0.3?

10.19 Dadas duas amostras $(X_1, \cdots, X_n)$, $(Y_1, \cdots, Y_n)$ ambas provenientes de uma mesma população $N(\mu, 1)$, qual é a distribuição de:

(a) Médias amostrais, $\overline{X}$ e $\overline{Y}$.

(b) Diferença de médias $\overline{X} - \overline{Y}$.

(c) Soma de médias $\overline{X} + \overline{Y}$.

(d) Média das médias $(\overline{X} + \overline{Y})/2$.

10.20 Um corredor procura controlar seus passos em uma corrida de 100 metros. Seus passos distribuem-se normalmente com média 0.97 metros e desvio padrão 0.1 metro. Determine a probabilidade de que 100 passos difiram de 100 metros por não mais de 5 metros.

10.21 Arredondam-se 20 números para o inteiro mais próximo e somam-se os números resultantes. Suponha que os erros individuais de arredondamentos são independentes e se distribuem uniformemente em $(-1/2, 1/2)$. Determine a probabilidade de que a soma obtida difira da soma dos vinte números originais por mais do que 3.

10.22 Fregueses chegam em certo supermercado segundo um processo de Poisson com intensidade média de dez por minuto. Seja $T_1, T_2, \cdots$ os tempos entre chegadas de fregueses, de modo que $T_1 + \cdots + T_n$ é o tempo de chegada do n-ésimo freguês.

(a) Utilizando o Teorema Central do Limite, ache um número entre 0 e 1 que seja aproximadamente igual à probabilidade do milésimo freguês chegar depois de 100 minutos.

(b) Como você calcularia o valor *exato* da probabilidade no item (a)? (Não se aceita uma integral no espaço de dimensão 1000).

10.23 Seja uma sequência de variáveis aleatórias $X_1, X_2, \cdots, X_n$ uniformemente distribuídas no interval (0,1), (0,2), etc. O que acontece com sua média aritmética quando n cresce?

10.24 Variáveis aleatórias $X_1, X_2, \cdots, X_n$ são uniformemente distribuídas, respectivamente, nos intervalos (-1,1), (-2,2), etc. Verifique se a média aritmética das variáveis aleatórias converge em probabilidade a zero quando n cresce.

10.25 Suponha que 30 dispositivos eletrônicos $D_1, D_2, \cdots, D_{30}$ estejam empregados da seguinte maneira: tão logo D_1 falhe, D_2 entra em operação; quando D_2 falha, D_3 entrará em operação, etc. Suponha que a duração até falhar D_i seja uma variável aleatória exponencialmente distribuída com parâmetro $\beta = 0.1\text{hora}^{-1}$. Seja T o tempo total da operação dos 30 dispositivos. Qual é a probabilidade de que T ultrapasse 350 horas?

10.26 Um computador, ao adicionar números arredonda cada número para o inteiro mais próximo. Admita-se que todos os erros de arredondamento sejam independentes e uniformemente distribuídos sobre (-0.5,0.5).

(a) Se 1500 números forem adicionados, qual é a probabilidade de que a magnitude do erro total ultrapasse 15?

(b) Quantos números poderão ser adicionados juntos de modo que a magnitude do erro total seja menor do que 10, com probabilidade 0.90?

10.27 Suponha que X_i, $i = 1, 2, \cdots, 50$ sejam variáveis aleatórias independentes, cada uma com distribuição de Poisson de parâmetro $\beta = 0.03$. Faça $S = X_1 + \cdots + X_{50}$.

(a) Empregando o Teorema Central do Limite, calcule $P(S \geq 3)$.

(b) Compare a resposta do item anterior com o valor exato dessa probabilidade.

10.28 A distribuição dos comprimentos dos elos da corrente de uma bicicleta tem distribuição normal com média 2cm e variância 0.01cm^2. Para que uma corrente se ajuste à bicicleta, deve ter comprimento total entre 58 e 61cm. Qual a probabilidade de que uma corrente com 30 elos não se ajuste à bicicleta? E com 29 elos?

Capítulo 11

Introdução aos Processos Estocásticos

11.1 Introdução

A modelagem analítica de qualquer fenômeno do mundo físico requer que simplificações sejam consideradas uma vez que o mundo físico é mais rico do que toda teoria desenvolvida em ciência e tecnologia. Em particular, se a modelagem analítica é estocástica inúmeras dependências têm de ser consideradas. Por exemplo, as variáveis aleatórias tempo entre mensagens recebidas, T_m, tempo de resposta em um sistema *on-line*, T_r, número de *jobs* esperando para serem processados, N_t, dependem do instante em que são consideradas; as compras realizadas na próxima semana em um supermercado podem depender da satisfação das compras realizadas até o presente momento; o estoque armazenado amanhã dependerá do nível de estoque de hoje bem como da demanda; o número de clientes esperando numa fila em uma certa hora depende do número de clientes que estavam na fila na hora anterior.

Um grande problema é que apesar dos modelos serem mais apropriados quando dependências são incluídas, estas por suas vez tornam os cálculos de probabilidades muito complicados ou impossíveis. Ou seja, quanto mais independência se assume no modelo probabilístico, maior a chance de poder realizar cálculos explicitamente o que implica na qualidade do modelo, onde qualidade do modelo significa a adequação do modelo proposto à realidade física [1]. Portanto, na modelagem analítica por meio de um modelo estocástico, o desafio é utilizar dependências que permitam que o modelo seja o mais fiel possível à realidade, mas que também sejam matematicamente tratáveis.

No caso das variáveis anteriormente mencionadas T_m, T_r e N_t é fácil supor que todas dependem do tempo em que são observadas. Na verdade então tem-se uma família de variáveis aleatórias, por exemplo, $\{N_t, t \in T\}$ onde T é o conjunto de todos os tempos durante o dia no qual o centro de computação está operando.

Definição 11.1.1: Uma família de variáveis aleatórias $\{X_t, t \in T\}$ é chamada de um *processo estocástico*.

$\forall t \in T$, X_t é uma variável aleatória. Fixado $t \in T$, o conjunto de todos os valores que X_t pode assumir é chamado de *espaço de estados*, Ω, e cada um desses valores é chamado

[1] A ciência e tecnologia desconhecem os adjetivos, consideram apenas os substantivos. Um modelo é bom quando se adequa ao mundo físico dentro de um critério topológico, por exemplo, uma distância.

de *estado do processo*. O espaço de estados corresponde ao espaço amostral e os estados do processo são os valores que, fixado t, a variável X_t assume. T é um conjunto de índices usualmente referenciado como *espaço de parâmetros*.

11.2 Classificação dos processos estocásticos

A *classificação* dos processos estocásticos é feita considerando-se o espaço de estados e o espaço de parâmetros:

(i) Espaço de parâmetros discreto. $T = \{0, 1, 2, \ldots\}$ ou $T = \{0, \pm 1, \pm 2, \ldots\}$.

(ii) Espaço de parâmetros contínuo. $T = \{t : -\infty < t < +\infty\}$ ou $T = \{t : t \geq 0\}$.

(iii) Espaço de estados discreto. Se fixado $t \in T$, a variável $X(t)$ assume um número finito ou enumerável de valores.

(iv) Espaço de estados contínuo. Se fixado $t \in T$, a variável $X(t)$ assume valores num conjunto não-enumerável.

Exemplo 11.2.1: Tempo de espera de uma consulta a uma base de dados, $\{W(t), t \geq 0\}$.

espaço de parâmetros: contínuo;
espaço de estados: contínuo.

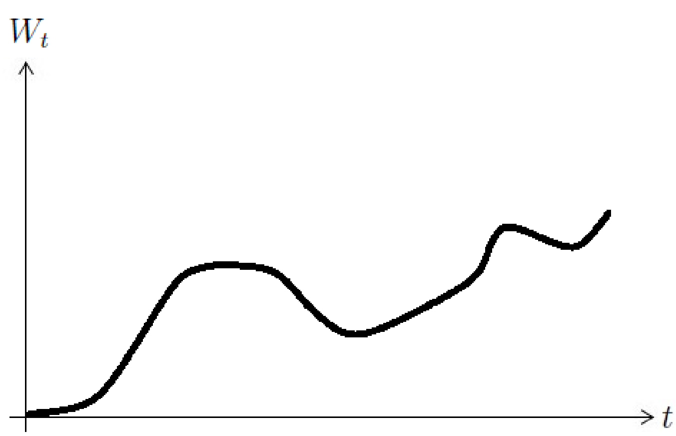

Figura 11.1: Representação de um processo estocástico com espaço de parâmetros e estados contínuo

■

Exemplo 11.2.2: Número de mensagens que chegam a determinado servidor no período de 0 a t, $\{N(t), t \geq 0\}$.

espaço de parâmetros: contínuo;
espaço de estados: discreto.

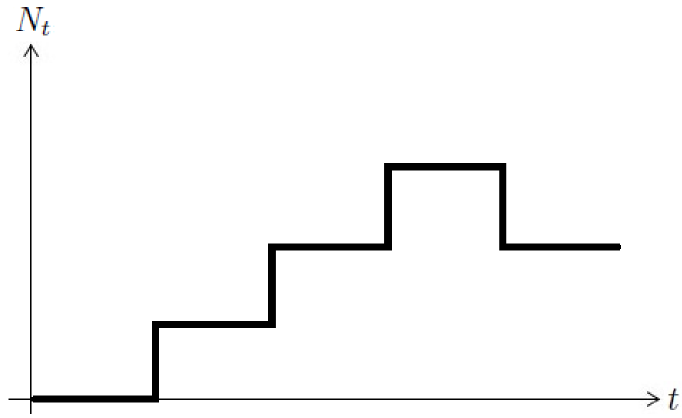

Figura 11.2: Representação de um processo estocástico com espaço de parâmetros contínuo e espaço de estados discreto

■

Exemplo 11.2.3: Tempo médio de execução de um *job* no n-ésimo dia da semana, $\{X_n, n = 1, \ldots, 7.\}$

espaço de parâmetros: discreto;
espaço de estados: contínuo.

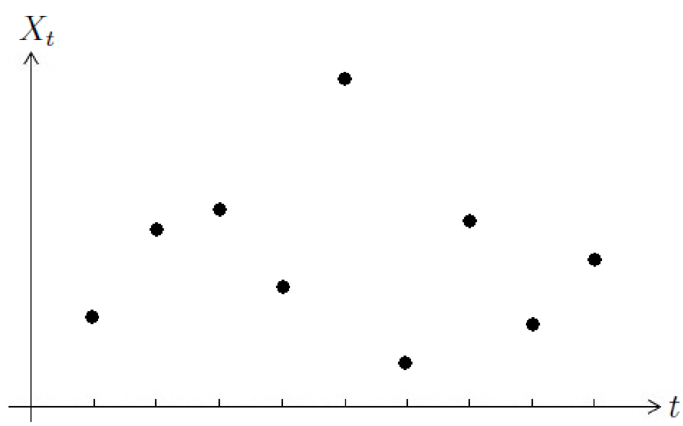

Figura 11.3: Representação de um processo estocástico com espaço de parâmetros discreto e espaço de estados contínuo

■

Exemplo 11.2.4: Número de *jobs* processados em um centro de processamento de dados no n-ésimo dia do ano, $\{X_n, n = 1, \ldots, 365.\}$

espaço de parâmetros: discreto;
espaço de estados: discreto.

■

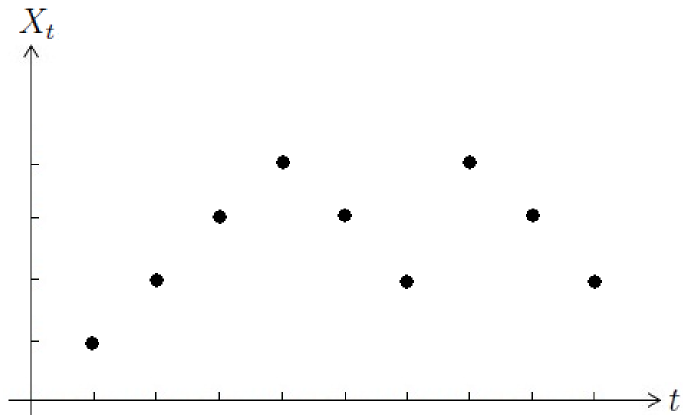

Figura 11.4: Representação de um processo estocástico com espaço de parâmetros e estados discreto

11.3 Exemplos de processos estocásticos

Alguns processos estocásticos são tão comuns no mundo físico que recebem atenção especial.

11.3.1 Processo de contagem

Sejam os eventos:

- uma chamada telefônica numa central de reservas de uma companhia aérea,

- a ocorrência de uma falha por *software* ou *hardware* em um sistema.

Tais eventos podem ser descritos por uma função de contagem, $N(t)$, definida para todo $t > 0$ como *o número de eventos que ocorreram depois do tempo 0 mas não mais tarde que o tempo t.*

Definição 11.3.1: $\{N(t), t \in T\}$ constitue um processo de contagem se:

(i) $N(0) = 0$.

(ii) $N(t)$ assume somente valores inteiros não-negativos.

(iii) $s < t \Rightarrow N(s) \leq N(t)$.

(iv) $N(t) - N(s)$ é o número de eventos que ocorreram em $[s, t)$.

11.3.2 Processo de Poisson

O processo de Poisson, com espaço de estados discreto e espaço de parâmetros contínuo, é útil em muitas aplicações práticas. O interesse é contar o número de eventos que ocorrem no intervalo de tempo $(0, t]$.

A próxima definição formaliza a ideia de "uma quantidade pequena com respeito a outra". A vantagem é que torna possível indicar esse fato sem especificar o exato relacionamento entre as duas quantidades.

Definição 11.3.2: f é $o(h)$ se $\lim_{h\to\infty} \frac{f(h)}{h} = 0$.

Definição 11.3.3: Um processo estocástico $\{X(t), t \geq 0\}$ com parâmetro contínuo tem

- *incrementos independentes* se eventos ocorrendo em intervalos de tempo não superpostos são independentes. Isto é, se $(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)$ não se sobrepõem então $X(b_1) - X(a_1), \ldots, X(b_n) - X(a_n)$ são variáveis aleatórias independentes.

- *incrementos estacionários* se $X(t+h) - X(s+h)$ tem a mesma distribuição que $X(t) - X(s)$ para cada escolha de s e t, $s < t$ e $h > 0$. Isto é, a distribuição de $X(t) - X(s)$ depende apenas do comprimento do intervalo e não dos possíveis valores s e t.

Definição 11.3.4: Um processo de contagem $\{N(t), t \geq 0\}$ é um processo de Poisson com taxa $\lambda > 0$ se:

(i) o processo tem incrementos independentes;

(ii) o processo tem incrementos estacionários;

(iii) $P(N(h) = 1) = \lambda h + o(h)$; (probabilidade da ocorrência de exatamente 1 evento em qualquer intervalo de tempo de comprimento h)

(iv) $P(N(h) \geq 2) = o(h)$; (probabilidade da ocorrência de mais do que 1 evento em qualquer intervalo de tempo de comprimento h)

Teorema 11.3.5: *Seja $\{N(t), t \geq 0\}$ um processo de Poisson com taxa $\lambda > 0$. Então a variável aleatória, Y, que conta o número de eventos em qualquer intervalo de tempo de comprimento $t > 0$ tem uma distribuição de Poisson com parâmetro λt,*

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

Portanto, o número médio de eventos ocorrendo em qualquer intervalo de tempo de comprimento t é λt e o número médio de eventos ocorrendo por unidade de tempo é $\frac{\lambda t}{t} = \lambda$.

O próximo teorema mostra outro importante atributo do processo de Poisson, que é a sua relação com a distribuição exponencial.

Teorema 11.3.6: *Seja $\{N(t), t \geq 0\}$ um processo de Poisson com taxa $\lambda > 0$. Sejam $0 < t_1 < t_2 < \ldots$ tempos sucessivos de ocorrências de eventos e $\{\tau_n, n = 1, 2, \ldots\}$ os tempos entre as ocorrências de eventos* [2]. *Então, $\{\tau_n\}$ são variáveis aleatórias independentes e identicamente distribuídas segundo uma exponencial com média $\frac{1}{\lambda}$.*

[2] $\tau_1 = t_1, \tau_2 = t_2 - t_1, \ldots$

O próximo teorema mostra que vale a recíproca do último teorema.

Teorema 11.3.7: *Seja $\{N(t), t \geq 0\}$ um processo de Poisson com taxa $\lambda > 0$ tal que os tempos entre chegadas de eventos, $\{\tau_n, n = 1, 2, \ldots\}$, sejam variáveis aleatórias independentes e identicamente distribuídas segundo uma exponencial com média $\frac{1}{\lambda}$. Então $\{N(t), t \geq 0\}$ segue um processo de Poisson com taxa λ.*

O processo de Poisson pode ser visto como um caso especial de um tipo mais geral de processo, que é importante para Teoria das Filas, o qual será visto a seguir.

11.3.3 Processo de nascimento-e-morte

A ideia intuitiva atrás de um processo de nascimento-e-morte é a de algum tipo de população que está ganhando novos membros através de nascimentos e perdendo velhos membros através de mortes. A população, para a maior parte das aplicações desses processos em computação é a dos usuários em uma fila de espera. A palavra usuários é genérica. Os usuários que chegam correspondem aos nascimentos e os que partem, após serem atendidos, às mortes.

Definição 11.3.8: Seja um processo estocástico com parâmetro contínuo $\{X_t, t \geq 0\}$ e com espaço de estados discreto $0, 1, \ldots$. Suponha que este processo descreve um sistema que está no estado E_n, $n = 0, 1, \ldots$ no tempo t se e só se $X(t) = n$ [3]. Então, o processo é descrito por um processo de nascimento-e-morte se existem *taxas de nascimento* $\{\lambda_n, n = 0, 1, \ldots\}$ e *taxas de morte* $\{\mu_n, n = 0, 1, \ldots\}$ tal que os seguintes postulados são satisfeitos:

(i) As únicas mudanças de estado são:

de E_n para E_{n+1}, $n \geq 1$;

de E_n para E_{n-1}, $n \geq 1$;

de E_0 para E_1.

Somente um nascimento ou uma morte ocorre no tempo t e não ocorre morte se o sistema está vazio.

(ii) Se no tempo t o sistema está no estado E_n, a probabilidade de que entre $(t, t+h)$ E_n mude [4] para E_{n+1} é $\lambda_n + o(h)$ e E_n mude para E_{n-1} é $\mu_n + o(h)$.

(iii) A probabilidade de que ocorra mais de uma transição $(t, t+h)$ é $o(h)$ [5].

Quando descreve-se um sistema de filas como um processo de nascimento-e-morte, o estado E_n corresponde aos n usuários no sistema esperando ou recebendo serviço.

[3] O sistema tem uma população de n elementos ou usuários no tempo t.

[4] Tem-se aqui probabilidades de transição, isto é de nascimento ou morte, em um pequeno intervalo de tempo quando o tamanho da população é n

[5] A probabilidade de mais de um nascimento ou mais de uma morte num intervalo de tempo muito pequeno é desprezível.

11.3.4 Processo de Markov

Definição 11.3.9: Um processo estocástico $\{X(t), t \in T\}$ é um processo de Markov se para qualquer conjunto $t_1 < t_2 < \ldots t_{n+1}$ no espaço de parâmetros T e $\{x_1, \ldots, x_{n+1}\}$ no espaço de estados Ω tem-se que

$$P(X(t_{n+1}) = x_{n+1} \mid X(t_1) = x_1, \ldots X(t_n) = x_n) = P(X(t_{n+1}) = x_{n+1} \mid X(t_n) = x_n).$$

Intuitivamente, esta definição significa que o futuro do processo depende somente do estado presente e do estado imediatamente anterior, e não de toda sua história.

Usualmente os processos de Markov são classificados como:

Cadeia de Markov com tempo discreto
Espaço de parâmetros: discreto. Espaço de estados: discreto

Processo de Markov com tempo discreto
Espaço de parâmetros: discreto Espaço de estados: contínuo

Cadeia de Markov com tempo contínuo
Espaço de parâmetros: contínuo Espaço de estados: discreto

Processo de Markov com tempo contínuo
Espaço de parâmetros: contínuo Espaço de estados: contínuo

11.4 Cadeias de Markov

Uma *Cadeia de Markov* é um processo de Markov com espaço de estados discreto.

Desde que uma cadeia de Markov tem espaço de estados discreto, os estados são rotulados por $\{E_0, E_1, \ldots\}$ ou, por conveniência notacional, $\{0, 1, \ldots\}$.

Para uma cadeia de Markov em tempo discreto é conveniente pensar que o processo faz **transições de estado** nos tempos t_n, $n = 0, 1, 2, \ldots$. Esta cadeia começa em um estado inicial i, quando $t = t_0$, isto é, $X_0 = i$ e faz uma transição de estado no próximo passo, $X_1 = j$.

11.4.1 Probabilidades de transição em um passo e distribuição inicial

Definição 11.4.1: As **probabilidades de transição em um passo** são definidas por

$$p_{ij}^{(1)} = p_{ij} = P(X_{n+1} = j \mid X_n = i),\ n, i, j = 0, 1, \ldots$$

e geralmente dependem de n.

Cadeias de Markov cujas probabilidades de transição não dependem de n são ditas como tendo **probabilidades de transição estacionárias** ou que são **homogêneas no tempo**, e neste caso,

$$P(X_{n+1} = j \mid X_n = i) = P(X_1 = j \mid X_0 = i), \forall n \geq 1.$$

As probabilidades de transição em um passo podem ser exibidas na forma de uma matriz quadrada chamada **matriz de probabilidades de transição**, P, da cadeia, como a seguir

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ p_{i0} & p_{i1} & p_{i2} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

Nesta matriz, as linhas correspondem aos estados do processo, isto é, aos valores que as variáveis aleatórias, $X_0, X_1, \ldots$ assumem para cada t fixo, $t = 0, t = 1, \ldots$ portanto tem-se que

$$p_{ij} \geq 0, \forall i, j = 0, 1, \ldots$$

e

$$\sum_{j=0}^{\infty} p_{ij} = 1, i = 0, 1, \ldots$$

As probabilidades de transição de estado em um passo podem ser representadas por um diagrama, chamado *diagrama de transição de estados*, como a seguir.

GRAFICO

Seja $p_i^{(n)}$ a probabilidade de que a cadeia de Markov discreta $\{X_n\}$ esteja no estado i no n-ésimo passo,

$$p_i^{(n)} = P(X_n = i), i = 0, 1, \ldots.$$

A **distribuição de probabilidade inicial** dos estados é

$$p_i^{(0)} = P(X_0 = i), i = 0, 1, \ldots.$$

Portanto,

$$0 \leq p_i^{(0)} \leq 1$$

e

$$\sum_{i=0}^{\infty} p_i^{(0)} = 1.$$

$p_i^{(0)}$ é comumente chamada de **distribuição inicial da cadeia**, ou, se representada na forma de um vetor, $(p_0^{(0)}, p_1^{(0)}, p_2^{(0)}, \ldots)$, denomina-se **vetor da distribuição inicial** .

Exemplo 11.4.2: Seja uma sequência de experimentos de Bernoulli onde a probabilidade de sucesso em cada experimento é $0 < p < 1$ e a de falha q, onde $p + q = 1$. Seja o estado do processo do processo na n-ésima repetição do experimento como sendo o número de sucessos ininterruptos até este instante. Então, X_n é o número de sucessos ininterruptos na n-ésima repetição do experimento e portanto o espaço de estados e o espaço de parâmetro são ambos $\{0, 1, \ldots\}$. Uma possível realização do experimento é $SFSSF$ e, neste caso, $n = 5$, $X_0 = 1$, $X_1 = 0$, $X_2 = 1$, $X_3 = 2$, $X_4 = 0$.

Calculando algumas probabilidades de transição em um passo, $p_{ij} = P(X_{n+1} = j \mid X_n = i)$:

$$\begin{aligned} p_{00} = P(X_{n+1} = 0 \mid X_n = 0) &= P(X_1 = 0 \mid X_0 = 0) \\ &= P(X_2 = 0 \mid X_1 = 0) \\ &= P(X_3 = 0 \mid X_2 = 0) = \ldots = q. \end{aligned}$$

$$\begin{aligned} p_{01} = P(X_{n+1} = 1 \mid X_n = 0) &= P(X_1 = 1 \mid X_0 = 0) \\ &= P(X_2 = 1 \mid X_1 = 0) \\ &= P(X_3 = 1 \mid X_2 = 0) = \ldots = p. \end{aligned}$$

$$\begin{aligned} p_{10} = P(X_{n+1} = 0 \mid X_n = 1) &= P(X_1 = 0 \mid X_0 = 1) \\ &= P(X_2 = 0 \mid X_1 = 1) \\ &= P(X_3 = 0 \mid X_2 = 1) = \ldots = q. \end{aligned}$$

$$\begin{aligned} p_{11} = P(X_{n+1} = 1 \mid X_n = 1) &= P(X_1 = 1 \mid X_0 = 1) \\ &= P(X_2 = 1 \mid X_1 = 1) \\ &= P(X_3 = 1 \mid X_2 = 1) = \ldots = 0. \end{aligned}$$

A matriz de probabilidades de transição para este processo é

$$P = \begin{pmatrix} q & p & 0 & 0 & \ldots \\ q & 0 & p & 0 & \ldots \\ q & 0 & 0 & p & \ldots \\ \ldots & \ldots & \ldots & \ldots & \end{pmatrix}$$

As probabilidades de transição independem de n, portanto tem-se uma cadeia de Markov homogênea no tempo. ■

11.4.2 Probabilidades de transição em n passos e distribuição de $X_n, n \geq 1$

Agora serão definidas as probabilidades de transição em n passos, ou seja, as probabilidades de que o processo estando no estado i no instante m estarão no estado j depois de n transições, isto é,

$$p_{ij}^{(n)} = P(X_{m+n} = j | X_m = i)$$

para $n \geq 0$ e $i, j \geq 0$.

Quando $n = 1$, $p^{(n)}(i,j) = p(i,j)$. Usando a definição de probabilidade condicional e o Teorema da probabilidade total[6], tem-se que

[6]Resultados do Capítulo 3

$$
\begin{aligned}
p_{ij}^2 = P(X_2 = j \mid X_0 = i) &= \frac{P(X_2 = j, X_0 = i)}{P(X_0 = i)} \\
&= \frac{\sum_z P(X_2 = j, X_1 = z, X_0 = i)}{P(X_0 = i)} \\
&= \frac{\sum_z P(X_0 = i) P(X_1 = z \mid X_0 = i) P(X_2 = j \mid X_1 = z)}{P(X_0 = i)} \\
&= \sum_z p(i,z)p(z,j).
\end{aligned}
$$

As equações de Chapman-Kolmogorov servem como um método para calcular as probabilidades de transição de n passos e são estabelecidas observando que

$$
\begin{aligned}
p_{ij}^{(m+n)} &= P(X_{m+n} = j | X_0 = i) = \sum_z P(X_n = z \mid X_0 = i) P(X_{m+n} = j \mid X_0 = i, X_n = z) \\
&= \sum_z p_{iz}^{(n)} P(X_{m+n} = j \mid X_0 = i, X_n = z) \\
&= \sum_z p_{iz}^{(n)} p_{zj}^{(m)}.
\end{aligned}
$$

Para uma cadeia de Markov tendo um número finito de estados, seja $P^{(n)}$ a matriz de probabilidades de transição em n passos, então, $P^{(n+m)} = P^{(n)} \cdot P^{(m)}$, onde $P^{(0)} = I$. Portanto,

$$P^{(n)} = P \cdot P^{(n-1)} = P \cdot P \cdot P^{(n-2)} = \ldots = P^n$$

e $P^{(n)}$ pode ser calculada multiplicando a matriz P por ela mesma n vezes.

Exemplo 11.4.3: Considere um sistema de comunicação que transmite os dígitos 0 e 1 através de vários estágios. Em cada estágio, a probabilidade de que o mesmo dígito seja recebido ou transmitido para próximo estágio é 0.75. Qual é a probabilidade de que um 0 que entrou no primeiro estágio seja recebido como 0 no quinto estágio?

Solução. O objetivo é calcular p_{00}^4. A matriz de probabilidades de transição e sua quarta potência são, respectivamente

$$P = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$$

$$P^4 = \begin{pmatrix} 0.53125 & 0.46875 \\ 0.46875 & 0.53125 \end{pmatrix}$$

Portanto, $p_{00}^{(4)} = 0.53125$. ■

A distribuição de X_n,

$$p_j^{(n)} = P(X_n = j)$$

pode ser obtida de duas formas distintas:

(i)

$$\begin{aligned} P(X_n = j) &= \sum_{i\in\Omega} P(X_0 = i, X_n = j) \\ &= \sum_{i\in\Omega} P(X_0 = i) P(X_n = j \mid X_0 = i) \\ &= \sum_{i\in\Omega} p_i^{(0)} p_{ij}^{(n)}. \end{aligned}$$

Esta fórmula calcula a distribuição de X_n em função da distribuição inicial e das probabilidades em n passos.

(ii)

$$\begin{aligned} P(X_n = j) &= \sum_{i\in\Omega} P(X_{n-1} = i, X_n = j) \\ &= \sum_{i\in\Omega} P(X_{n-1} = i) P(X_n = j \mid X_{n-1} = i) \\ &= \sum_{i\in\Omega} P(X_{n-1} = i) p_{ij}^{(1)} \\ &= \sum_{i\in\Omega} p_i^{(n-1)} p_{ij}^{(1)}. \end{aligned}$$

11.4.3 Classificação dos estados e distribuições limite

Definição 11.4.4: Um estado i é **absorvente** se e só se $p_{ii} = 1$.

Definição 11.4.5: Sejam i e j dois estados não necessariamente distintos. Diz-se que i conduz a j ou j é acessível a partir de i se existe $n \geq 0$ tal que $p^{(n)}(ij) > 0$. Se todo estado é acessível a partir de algum outro, a cadeia é dita **irredutível**.

Definição 11.4.6: Sejam i e j dois estados não necessariamente distintos. Se $i \to j$ e $j \to i$ então i e j são **comunicantes**, isto é, $i \leftrightarrow j$. Neste caso, existem $n, m \geq 0$ tais que $p_{ij}^{(n)} > 0$ e $p_{ji}^{(m)} > 0$.

Definição 11.4.7: O estado i é **periódico** com período $d > 0$ se $p_{ii}^{(n)} > 0$ somente para $n = d, 2d, 3d, \ldots$ onde d é o maior inteiro que verifica a propriedade. Se $d = 1$, i é **aperiódico**.

Definição 11.4.8: Para cada i seja $f_i^{(n)}$ a probabilidade de que o primeiro retorno ao estado i ocorra exatamente em n passos (ou transições) depois de deixar i, isto é,

$$f_i^{(n)} = P(X_n = i, X_k \neq i \ \forall k = 1, \ldots, n-1 \mid X_0 = i)$$

e

$$f_i^{(0)} = 1, \ \forall i.$$

Seja f_i a probabilidade de retorno ao estado i, logo

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

Se

$$f_i < 1, \ i \text{ é um estado } \textbf{transiente},$$

Se

$$f_i = 1, \ i \text{ é um estado } \textbf{recorrente}.$$

Se i é recorrente, o **tempo médio de recorrência a** i é

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

Se

$$\mu_i = \infty, \ i \text{ é } \textbf{nulo recorrente},$$

Se

$$\mu_i < \infty, \ i \text{ é } \textbf{recorrente positivo}, \text{ ou } \textbf{recorrente não-nulo}.$$

Definição 11.4.9: Uma cadeia de Markov discreta é dita como tendo uma distribuição de probabilidade estacionária,

$$p = (p_1, p_2, \ldots)$$

onde $p_i \geq 0$ e $\sum_i p_i = 1$, se a equação matricial

$$p = pP$$

é satisfeita.

A equação matricial $p = pP$ pode ser escrita como o sistema de equações lineares

$$p_j = \sum_i p_j p_{ij}, j = 0, 1, \ldots$$

Definição 11.4.10: Uma cadeia de Markov tem distribuição de probabilidade limite

$$p = (p_1, p_2, \ldots)$$

se

$$\lim_{n\to\infty} p_j^{(n)} = \lim_{n\to\infty} P(X_n = j) = p_j, j = 0, 1, \ldots$$

Definição 11.4.11: Uma cadeia de Markov discreta, irredutível, aperiódica e onde todos os estados sejam recorrentes positivos é denominada de **ergódica**.

Para uma cadeia de Markov ergódica a distribuição de probabilidade estacioária e a distribuição limite são iguais, sendo chamadas distribuições de equilíbrio ou estáveis.

Exemplo 11.4.12: No Exemplo 11.4.3 visto, como a cadeia de Markov é irredutível e aperiódica é também ergódica. Portanto, será calculada a distribuição de probabilidade de equilíbrio.

$$\sum_i p_i = 1,$$

$$p_j = \sum_i p_i P_{ij}, j = 0, 1, \ldots$$

Logo,

$$\begin{cases} p_0 & + & p_1 & & = & 1 \\ p_0 & = & 0.75p_0 & + & 0.25p_1 \\ p_1 & = & 0.25p_0 & + & 0.75p_1 \end{cases}$$

A solução deste sistema é $p_0 = p_1 = 0.5$. A interpretação do resultado obtido é que se os dados passam através de muitos estágios, a saída independe da entrada original e cada dígito recebido é igualmente provavél de ser 0 ou 1. ■

Exemplo 11.4.13: Todo ano um homem troca seu carro por um carro novo. Se tem um Palio ele o troca por um Gol. Se tem um Gol ele o troca por um sedan. Contudo, se tem um sedan, é tão provável que o troque por um novo sedan quanto por um Palio ou Gol. Em 2005 ele comprou seu primeiro carro, que foi um sedan. Calcule a probabilidade de que

(a) tenha um sedan em 2007;

(b) tenha um Palio em 2007;

(c) tenha um Gol em 2008;

(d) após um número suficientemente grande de anos, com que frequência terá tido um sedan?

Solução. O espaço de estados é

$$\Omega = \{\text{Palio, Gol, sedan}\}.$$

O espaço de parâmetros é

$$T = \{0, 1, 2, \ldots\}$$

O vetor de probabilidades iniciais é

$$p^{(0)} = (0, 0, 1).$$

A matriz de probabilidades de transição é

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

$$p^{(2)} = p^{(0)}P^2 = (\frac{1}{9}, \frac{4}{9}, \frac{4}{9}).$$

(a) $p_3^{(2)} = \frac{4}{9}$.

(b) $p_1^{(2)} = \frac{1}{9}$.

(c) $p^{(3)} = p^{(2)}P = (\frac{4}{27}, \frac{7}{27}, \frac{18}{27})$. Logo, $p_2^{(3)} = \frac{7}{27}$.

(d) Encontrando o vetor fixo de probabilidades:

$$(x, y, z)P = (x, y, z) \Rightarrow (x, y, z) = (\frac{1}{6}, \frac{2}{6}, \frac{3}{6}).$$

Logo, 50% das vezes terá tido um sedan.

■

11.5 Exercícios

11.1 Nas observações feitas por Rutherford e Geiger, uma substância radioativa emitia uma média de 3.87 partículas α durante 7.5 segundos.

(a) Qual é a probabilidade de que a substância emita pelo menos uma partícula α por segundo?

(b) Neste caso, isto é, considerando o número de partículas α emitidas por segundo, escreva a expressão da densidade da variável aleatória correspondente ao tempo entre chegadas.

11.2 **(a)** Determine a distribuição de probabilidade do intervalo de tempo entre chegadas sucessivas, T, num processo de Poisson de parâmetro βt. Sugestão: $\{T < t\} \Leftrightarrow \{X_t > 0\}$.

(b) Considere um sistema computacional onde o fluxo de chegadas de um especifífico programa por hora, X_t, segue um processo de Poisson com taxa média de 60 *jobs*. Determine a probabilidade de que o intervalo de tempo entre *jobs* sucessivos seja menor que 8 minutos.

11.3 Prove que a distribuição dos intervalos de tempo entre sucessivos eventos num processo de Poisson com intensidade λ, é uma exponencial com parâmetro λ.

11.4 A rã Dõ descansa sobre o vértice A de um triângulo ABC. A cada minuto a rã salta do vértice em que está para um vértice adjacente com probabilidade $p \in (0, 1)$ do salto ser no sentido horário e $1 - p$ do salto ser no sentido anti-horário. Modele o problema por uma Cadeia de Markov.

(a) Especifique o espaço de estados, E.

(b) Especifique o espaço de parâmetro, T.

(c) Especifique a matriz de probabilidade de transição, P.

(d) A distribuição de probabilidade inicial, $p^{(0)}$.

(e) Verifique (fácil e rapidamente) que P é *regular*.

(f) Encontre a distribuição de probabilidade estacionária da cadeia.

(g) Considere que o triângulo é equilátero e atribua um valor numérico conveniente para p. No terceiro minuto, com que probabilidade a rã Dõ poderá estar em cada um dos vértices do triângulo ABC? Especifique claramente sua resposta.

11.5 Três crianças A, B e C estão arremessando uma bola uma para a outra. A sempre arremessa a bola para B e B sempre arremessa a bola para C. É tão provável que C lance a bola para B quanto para A. Seja $X_n,\ n = 1 \ldots n$ a nésima pessoa a arremessar a bola.

(a) Especifique o espaço de estados do processo.

(b) Especifique a *matriz de probabilidades de transição*, M.

(c) Esboçe o *diagrama de transição de estados*.

(d) Considerando M, M^2 e M^3, você diria que M é regular? Justifique sua resposta.

(e) Determine o único vetor fixo de probabilidades. Interprete o resultado.

11.6 Os hábitos de estudo de um estudante são os seguintes: se estuda uma noite, tem 70% de certeza que não estudará na noite seguinte; em contrapartida, se não estuda uma noite, tem 60% de certeza de que não estudará também na noite seguinte. Com que freqüência ele estuda numa seqüência suficientemente grande de dias?

11.7 Suponha que o estado social, ou financeiro, de uma pessoa possa ser classificado como de *classe baixa*, denotado por 1, de *classe média*, denotado por 2 ou de *classe alta*, denotado por 3. Admite-se que a classe de um indivíduo depende apenas da classe do seu predecessor imediato. Seja a matriz de probabilidades de transição dada por

$$P = \begin{pmatrix} 0.6 & 0.3 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.2 & 0.7 \end{pmatrix}$$

e, que, em dada geração inicial, o percentual de indivíduos nas classes 1, 2 e 3 é, respectivamente, 40%, 50% e 10%.

(a) Se um indivíduo é da *classe média*, qual é a probabilidade de que seu neto venha a ser da *classe alta*?

(b) Quais as percentagens de indivíduos em cada classe daqui a três gerações?

(c) Encontre a distribuição de equilíbrio da cadeia e a interprete.

11.8 Suponha que determinado produto seja fabricado pelas companhias A e B. Presentemente a companhia A desfruta de 75% do mercado e a B dos 25% restante. Uma pesquisa realizada revelou que, de uma ano para outro, 15% dos consumidores da companhia A passam a consumir o produto da companhia B e 10% dos consumidores da companhia B, passam-se para a companhia A. Sabe-se que, dada a companhia com que o consumidor comercia atualmente, a companhia com que ele estará comerciando no ano seguinte será independente das companhias com as quais comerciou no passado.

(a) Especifique os espaços de estados e de parâmetro da cadeia.

(b) Especifique a matriz de probabilidades de transição de probabilidades, P.

(c) Especifique a distribuição de probabilidade inicial da cadeia.

(d) Que percentagem do mercado caberá à companhia B daqui a 4 anos?

(e) P é regular? P tem algum estado absorvente? Justifique sua resposta.

(f) Sendo possível, justifique a razão, determine a distribuição de equilíbrio da cadeia de Markov. Interprete, com o mínimo de palavras possível, o resultado obtido.

11.9 Uma dada impressora tem, sistematicamente, apresentado uma das seguintes situações:

c: está funcionando corretamente;
d: está apresentando algum tipo de defeito;
n: não está funcionando.

O pessoal do suporte decide observá-la todos os dias às 8:00. No dia inicial do começo das observações a impressora está funcionando corretamente. A experiência tem mostrado que o funcionamento da impressora pode ser modelado por uma cadeia de Markov com a seguinte matriz de probabilidades de transição:

$$\begin{pmatrix} 0.3 & 0.5 & 0.2 \\ 0 & 0.2 & 0.8 \\ 0 & 0 & 1 \end{pmatrix}$$

(a) Esta matriz apresenta algum estado absorvente? Se sim, qual, justificando sua resposta.

(b) Como serão as probabilidades de transição no terceiro dia de observação? Qual o valor de $p_{cn}^{(3)}$?

(c) No terceiro dia de observação, qual o estado mais provável da impressora?

(d) Seria possível estudar o comportamento desta impressora após um número grande de dias de observação? Justifique sua resposta.

11.10 Seja

$$A = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

uma matriz estocástica e $u = (u_1, u_2, u_3)$ um vetor de probabilidade.

(a) Mostre que uA também é um vetor de probabilidade.

(b) Suponha que u seja um vetor fixo de probabilidade. Mostre que ku, $k > 0$ também é um vetor fixo de probabilidade..

11.11 O território de um vendedor é constituído de três cidades, A, B e C. Ele nunca vende na mesma cidade em dias consecutivos. Se vende na cidade A, no dia seguinte vende na cidade B. Contudo, se vende na B ou em C, então no dia seguinte é duas vezes mais provável que ele venda em A do que na outra cidade.

(a) Especifique o espaço de estados do processo.

(b) Especifique a matriz de probabilidade de transição M.

(c) Esta cadeia tem algum estado absorvente?

(d) Após um número sufucientemente grande de dias, com que freqüência ele vende em cada uma das cidades?

Capítulo 12

Análise Exploratória de Dados

12.1 Tipos de Variáveis

Quando é necessário analisar um conjunto de dados decorrentes da realização de um experimento, ou da observação do mundo real, é comum a aplicação de um conjunto de técnicas que servem como um indicativo de qual procedimento deve ser adotado. Estas técnicas quando corretamente aplicadas e interpretadas fornecem valioso suporte para a tomada de decisões quer com respeito aos dados em si, quer com respeito a qual método de inferência aplicar.

Nem todas as variáveis são numéricas ou **quantitativas**, como as que foram estudadas nos capítulos anteriores, pode-se ter uma variável não numérica ou **qualitativa** como, por exemplo, modelos distintos de sistemas operacionais, ou distintos paradigmas de liguagens de programação, ou diferentes classes da população com respeito ao número de salários mínimos ganhos. As variáveis quantitativas podem ser *discretas* ou *contínuas*, enquanto as qualitativas podem ser classificadas como *nominais* ou *ordinais*, dependendo se existe ou não uma ordem natural em seus possíveis resultados. Por exemplo, o tipo de sistema operacional é uma variável qualitativa nominal, enquanto a classe da população com respeito ao número de salários mínimos ganhos é uma variável qualitativa ordinal.

Um outro possível critério para classificar variáveis é em função da escala de medida adotada para se analisar o resultado do experimento. As escalas de medidas podem ser: *nominais*, *ordinais*, *intervalares*, e de *razão*.

Uma escala nominal é utilizada para classificar os resultados de um experimento, por exemplo, se dado equipamento falhou ou não durante o período de estudo, a marca e o modelo do equipamento em questão.

Uma escala ordinal além de classificar os resultados também pode ser utilizada para estabelecer uma ordem entre as diferentes classes de possíveis resultados, por exemplo, grau de escolaridade de um indivíduo, classe socio-econômica de um indivíduo, posição que um indivíduo conclui uma certa corrida. A estrutura da escala ordinal não é alterada por transformações estritamente monotônicas.

Uma escala intervalar pode ser utilizada para além de classificar e ordenar os resultados também quantificar diferença entre as classes. Nesta escala é necessário estabelecer uma origem arbitrária e uma unidade de medida. Por exemplo, a temperatura de um dado equipamento em funcionamento medida em graus centígrados constitui uma medida numa escala intervalar. Considere o caso em que tem-se três equipamentos E1, E2, e E3, operando em temperaturas de 40, 45 e 80 graus centígrados, respectivamente; é válido afirmar que a diferença de temperatura entre E3 e E2 é 7 vezes maior que a diferença de temperatura entre E2 e E1. Contudo, nesta escala não faz sentido afirmar que E3 tem uma temperatura 2 vezes maior que E1, pois a origem e a unidade de graus centígrados escolhidas são arbitrárias, se a temperatura estivesse sendo medida em graus Fahrenheits não se observaria esta relação.

Uma escala de razão podem ser utilizada para além de classificar e ordenar os resultados também estabelecer quão maior é um resultado que outro. A diferença com a escala intervalar é que agora existe um zero bem definido neste escala. A altura de um indivíduo, o tempo até ocorrência de um dado evento, o número de ocorrências de certo evento em um dado intervalo de tempo são exemplos de medidas que utilizam uma escala de razão. No caso de dois equipamentos E1 e E2 com tempo de vida útil de 100h e 200h, respectivamente, é válido afirmar que o tempo de vida útil de E2 é o dobro do tempo de vida útil de E1.

12.2 Análise preliminar de um conjunto de observações

O que será visto a seguir é como proceder uma análise preliminar em um conjunto de dados.

12.2.1 Representações Gráficas

Quando se procede uma análise em um conjunto de dados resultantes de variáveis quantitativas ou qualitativas o que é inicialmente feito é um gráfico. Existem vários tipos de gráficos, sendo os mais comumentes encontrados: barras, colunas, setores, pictograma, gantt, kiviat, linhas, histograma, box-plot e q-q plot. O tipo de gráfico a ser usado depende do tipo de variável do problema, se *qualitativa* ou *quantitativa*.

Para representar a distribuição dos dados de uma variável qualitativa, os dois gráficos mais utilizados são: o *gráfico de barras* e o *gráfico de setores* ou *pizza*. Esses gráficos consideram que o conjunto de dados pode ser particionado em classes.

O gráfico de barras consiste em construir retângulos ou barras paralelas, uma para cada classe, em que uma das dimensões é proporcional à frequência de ocorrência ou representatividade desta classe, e a outra dimensão é arbitrária porém igual para todas as barras. As barras podem ser horizontais ou verticais.

O gráfico de setores ou pizza consiste de um círculo de raio arbitrário, representando a totalidade dos dados, dividido em fatias sendo que cada fatia corresponde a uma classe e tem área proporcional à representatividade desta classe nos dados.

Para uma variável quantitativa discreta também utiliza-se um gráfico de barras como no caso de variáveis quantitativas, onde agora tem-se uma barra para cada possível valor que a variável pode assumir. Também considera-se um gráfico de dispersão unidimensional onde desenha-se apenas pontos no plano cartesiano da forma (x_i, n_i), isto é, onde a abscissa do ponto é um possível valor da variável e a ordenada é a frequência de ocorrência deste valor.

Uma outra alternativa de gráfico para variável quantitativa que é muito útil no caso de variáveis contínuas o *histograma*.

Para a construção de um histograma, o primeiro passo é definir os intervalos contíguos e disjuntos que cubram todos os resultados observados. Uma vez definidos os intervalos, um histograma nada mais é do que um gráfico de barras contíguas. Como é uma representação gráfica, a priori não é necessário rigor na construção dos retângulos, entretanto se o for a base é proporcional ao comprimento do intervalo e a área da barra é proporcional ao número de dados neste dado intervalo. Logo, se o i-ésimo intervalo tem comprimento Δ_i e a frequência relativa de ocorrência de resultados neste intervalo é f_i, então a altura da barra deve ser proporcional a f_i/Δ_i, que é chamada de *densidade de frequência* da i-ésima classe. Com essa convenção a área total do histograma é proporcional a 1.

12.2.2 Sumarizando Observações

Considere a TabelařefTabelaDados que contém informações sobre empregados de uma companhia.

Tabela 12.1: Dados sobre empregados de uma companhia.

No.	Estado Civil	Grau de Instrução	No. de Filhos	Salário	Idade	Sexo
1	S	Médio	0	3	34	F
2	C	Superior	2	5	25	M
3	C	Fundamental	1	4	46	M
4	C	Fundamental	3	5.5	32	M
5	S	Médio	1	7.3	23	F
6	C	Médio	2	3.5	39	M
7	S	Superior	3	10	50	M
8	C	Médio	4	6	47	M
9	C	Médio	0	2	21	F
10	S	Médio	1	3.7	33	M

Uma maneira útil de se descrever os resultados das variáveis é através da frequência (absoluta), frequência relativa (proporção) e porcentagem. Por exemplo, considerarando a variável Grau de Instrução na Tabela 12.1. A frequência de uma dada classe é o número de vezes que determinada classe ocorreu nos resultados do experimento. A frequência relativa é a proporção de vezes que dada classe ocorreu em relação ao número total de indivíuos que participaram do experimento. A porcentagem é igual a 100 vezes a frequência relativa. A Tabela 12.2 é conhecida como tabela de frequência para a variável Grau de Instrução.

Quando o objetivo for comparar esta variável Grau de Instrução entre diferentes empresas, deve-se usar ou a frequência relativa ou a porcentagem, pois possuem o mesmo total para qualquer empresa, enquanto o número total de empregados varia de empresa para empresa.

Tabela 12.2: Tabela de Frequências.

Grau de Instrução	Frequência (n_i)	Frequência Relativa (f_i)	Porcentagem $100f_i$
Fundamental	2	0.2	20
Médio	6	0.6	60
Superior	2	0.2	20
Total	10	1	100

Em geral, quando uma tabela de frequência é construída o objetivo é resumir os resultados no que diz respeito a uma dada classe. No caso de uma variável quantitativa, se faz necessário que dividam-se em intervalos os possíveis resultados do experimento para esta variável, pois caso contrário pode acontecer que cada resultado ocorra somente um número pequeno de vezes e não se possa resumir a informação a respeito da dada variável. Esta situação ocorre frequentemente no caso de variáveis que assumem valores reais. No exemplo anterior, suponha que se deseje construir uma tabela de frequência para a variável Salário. Neste caso, pode-se considerar intervalos de tamanho 3 para construir a seguinte tabela:

Tabela 12.3: Tabela de frequências para variável quantitativa.

Salário	Frequência (n_i)	Frequência Relativa (f_i)	Porcentagem $100f_i$
$[0, 3)$	1	0.1	10
$[3, 6)$	6	0.6	60
$[6, 9)$	2	0.2	20
$[9, 12)$	1	0.1	10
Total	10	1	100

A escolha dos intervalos acima é arbitrária, dependendo do contexto cada profissional pode escolher um conjunto diferente de intervalos. A única restrição que tal escolha deve satisfazer é que estes intervalo sejam disjuntos e que cubram todos os valores que foram obtidos pela variável no experimento. Se forem escolhidos poucos intervalos, perde-se informação, pois note que a Tabela 12.3 só afirma que 6 pessoas têm salário entre 3 e 6 salários mínimos sem especificar qual o salário exato deles. Por outro lado, se muitos intervalos são escolhidos, então a intenção de resumir os resultados do experimento não é cumprida. Em geral, recomenda-se o uso de 5 a 15 intervalos de comprimentos iguais.

Um lembrete: a Estatística é mais velha que o computador digital. O significado desta declaração é que os estatíticos do passado desenvolveram fórmulas fantásticas para calcular indicadores quando nem de calculadoras dispunham. Portanto, os indicadores estatísticos vistos a seguir podem ser computados ou *sem o uso de computadores* ou *com o uso de* *computadores*, isto é, ou são usadas fórmulas já completamente exauridas na literatura, ou quase nenhuma fórmula é usada.

Cenário: seja um conjunto de dados, ou observações,

$$x_1, x_2, \cdots, x_n$$

que podem ser de uma população, ou amostra, ou mais de um conjunto de dados, de uma variável *quantitativa*, *discreta* ou *contínua*. O objetivo, usualmente, é decidir algo a partir dos dados considerados.

Problema: Sumarizar informações sobre o conjunto de dados.

As medidas ou os indicadores estatísticos mais comumente usados são

(i) as medidas de *tendência central* ou **posição**: média aritmética ($\overline{x}$) , mediana ($\tilde{x}$) e moda ($\hat{x}$);

(ii) as medidas de **dispersão**: amplitude (r), variância (σ^2), desvio padrão (σ) e coeficiente de variação (cov);

(iii) as **separatrizes** ou **quantis**:

quartis,

$$q(p), p = \frac{1}{4}, \frac{2}{4}, \frac{3}{4},$$

decis,

$$q(p), p = \frac{1}{10}, \frac{2}{10}, \cdots, \frac{9}{10},$$

centis ou percentis,

$$q(p), p = \frac{1}{100}, \frac{2}{100}, \cdots, \frac{99}{100}.$$

A *média aritmética* de uma variável é a soma dos seus valores divididos pelo número total de resultados obtidos.

A *moda* de uma variável é definida como sendo o seu resultado que ocorreu com maior frequência durante o experimento. A moda não é necessariamente única. Se houver empate entre a frequência de ocorrência de mais de dois possíveis resultados, então todos serão moda da variável em questão. A moda não é necessariamente numérica, por exemplo, se a variável em questão for grau de instrução na população brasileira, a moda é onde tem uma maior concentração de indivíduos. Também uma variável pode não ter moda, o que acontece se não houver valores que se repitam.

A *mediana* é o resultado que ocupa a posição central de uma série de observações, quando estas estão ordenadas. Portanto, metade dos valores são menores e metade são maiores que a mediana. Note que a mediana também é uma separatriz, corresponde a $q(\frac{2}{4}) = q(\frac{5}{10}) = q(\frac{50}{100})$. Quando o número de observações for par a mediana é a média aritmética entre as duas observações centrais.

A presença de valores ou muito pequenos ou muito grandes alteram sua média e não alteram a mediana. Portanto, a mediana é muitas vezes usada para representar uma medida central de um conjunto de observações.

As medidas de posição vistas informam sobre a posição central dos resultados mas não sobre sua variabilidade. Para tanto são necessárias as medidas de dispersão. Por exemplo, considere dois grupos de resultados de uma certa variável: Grupo 1 = $\{3, 4, 5, 6, 7\}$ e Grupo 2 = $\{1, 3, 5, 7, 9\}$. Ambos os grupos possuem a mesma média e mediana que é igual a 5, porém os resultados do Grupo 1 estão mais aglutinados ao redor deste valor. Medidas de dispersão são utilizadas para mensurar esta variabilidade. Estas medidas analisam quão distante da média estão os resultados. Para uma variável, a medida de dispersão mais usada é o *desvio padrão*, e para mais de uma variável, o *coeficiente de variação*.

Apenas as medidas de posição e de dispersão não informam a respeito da simetria ou assimetria da distribuição dos resultados. Os quantis ou separatrizes são medidas que servem para informar a este respeito.

Um *quantil de ordem* p ou p*-quantil*, é indicado por $q(p)$, onde p é uma proporção qualquer, $0 < p < 1$, e é tal que $100p\%$ dos resultados sejam menores que $q(p)$. Existem alguns quantis que são usados mais frequentemente e recebem nomes particulares: $q(0.25)$ é o 1o. quartil ou 25o. percentil; $q(0.5)$ é a mediana, 5o. decil, ou 50o. percentil; $q(0.75)$ é o terceiro quartil ou 75o. percentil; $q(0.95)$ é o 95o. percentil.

Uma medida de dispersão é a *distância interquartil*, d_q, definida como sendo a diferença entre o terceiro e o primeiro quartil, isto é, $d_q = q(0.75) - q(0.25)$.

Os cinco valores $x_{(1)}$, $q(0.25)$, $q(0.5)$, $q(0.75)$, e $x_{(n)}$ são importantes para se ter uma idéia a respeito da assimetria da distribuição dos dados. Para se ter uma distribuição aproximadamente simétrica, é preciso que:

(a) $q(0.5) - x_{(1)} \simeq x_{(n)} - q(0.5)$;

(b) $q(0.5) - q(0.25) \simeq q(0.75) - q(0.5)$;

(c) $q(0.25) - x_{(1)} \simeq x_{(n)} - q(0.75)$.

A questão que se coloca agora é se os dados serão ou não agrupados. Até antes do computador, dependendo da quantidade de dados, agrupá-los nas chamadas *distribuições de frequências* era inevitável. Depois do computador, agrupar os dados só se for necessário observar algum tipo de padrão de comportamento funcional deles. A seguir será visto como calcular os indicadores estatísticos referenciados anteriormente, para dados agrupados e não agrupados. Neste livro as fórmulas para dados agrupados não serão provadas, uma vez que as mesmas encontram-se em vários dos textos básicos sobre o assunto.

12.2.3 Dados agrupados

Se os dados estão agrupados então estão dispostos numa tabela como a Tabela 12.4, a qual tem k *classes* com respectivas frequências absolutas n_i, $i = 1, \cdots, k$, limites inferiores l_i e superiores L_i. Usualmente o limite superior da classe i é igual ao inferior da classe $i+1$, isto é, $L_1 = l_2$, $L_2 = l_3$, etc.

Medidas de Tendência Central

Tabela 12.4: Tabela de frequências para dados agrupados.

Classes	n_i
$l_1 \vdash L_1$	n_1
$l_2 \vdash L_2$	n_2
$l_3 \vdash L_3$	n_3
$\cdots$	$\cdots$
$l_k \vdash L_k$	n_k
Total	n

Média aritmética, $\overline{x}$

$$\overline{x} = \frac{x_1 n_1 + \ldots + x_k n_k}{n_1 + \ldots + n_k} = \frac{1}{n} \sum_{i=1}^{k} x_i n_i,$$

onde $x_i = \frac{l_i + L_i}{2}$.

Mediana, $\tilde{x}$

A mediana é o valor que divide o conjunto de dados em duas partes iguais, isto é, 50% são menores ou iguais que $\tilde{x}$ e 50% são maiores ou iguais que $\tilde{x}$. A mediana não é única; seu cálculo implica no cálculo das frequências acumuladas Fc_i. Neste caso tem-se:

Frequências acumuladas

Classes	n_i	Fc_i
$l_1 \vdash L_1$	n_1	n_1
$l_2 \vdash L_2$	n_2	$n_1 + n_2$
$l_3 \vdash L_3$	n_3	$n_1 + n_2 + n_3$
$\cdots$	$\cdots$	$\cdots$
$l_k \vdash L_k$	n_k	$n_1 + n_2 + \cdots n_k$
Total	n	-

A fórmula da mediana é a seguinte.

$$\tilde{x} = l_i + (\frac{\frac{n}{2} - Fc_{i-1}}{n_i})(L_i - l_i),$$

onde,

L_i, limite superior da classe da mediana,
l_i, limite inferior da classe da mediana,
n_i, frequência absoluta da classe da mediana,
Fc_{i-1}, somatório das frequêncis anteriores à classe da mediana.

Moda, $\hat{x}$

A moda é o valor que mais se repete, portanto está associado à classe de maior frequência absoluta. Uma distribuição pode ter mais de uma moda.

$$\hat{x} = l_i + (\frac{\Delta_1}{\Delta_1 + \Delta_2})(L_i - l_i),$$

onde,

L_i, limite superior da classe modal,
l_i, limite inferior da classe modal,
$\Delta_1 = n_i - n_{i-1}$,
$\Delta_2 = n_i - n_{i+1}$,

sendo n_i a frequência absoluta da classe modal.

Medidas de Dispersão

Amplitude, r

$$r = x_{(n)} - x_{(1)},$$

onde $x_{(1)}$ e $x_{(n)}$ são, respectivamente, o menor e o maior valor do conjunto de observações.

Variância, σ^2

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x})^2 n_i = \frac{1}{n}\sum_{i=1}^{k} x_i^2 - (\overline{x})^2,$$

sendo $x_i = \frac{l_i + L_i}{2}$.

Desvio padrão, σ

$$\sigma = +\sqrt{\sigma^2}.$$

Coeficiente de variação, cov

$$cov = \frac{\sigma}{\overline{x}}.$$

Separatrizes

A fórmula da mediana, com as devidas modificações, pode ser usada para calcular qualquer separatriz: mediana, quantis, decis e centis (ou percentis).

O cálculo de qualquer quantil pode ser realizado através de uma modificação no numerador da fórmula da mediana da seguinte maneira. O termo $\frac{n}{2}$ pode ser lido como $\frac{1}{2}n$; essa fração indica qual a classe da mediana através da observação da frequência acumulada. Portanto, para o i-ésimo quartil, $i = 1, \cdots, 3$, j-ésimo decil, $j = 1, \cdots, 9$ ou k-ésimo centil, $k = 1, \cdots, 99$ tem-se, respectivamente, $\frac{i}{4}n$, $\frac{j}{10}n$ e $\frac{k}{100}n$.

12.2.4 Dados não-agrupados

Medidas de Tendência Central

Média aritmética, $\overline{x}$

$$\overline{x} = \frac{x_1 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Mediana, $\tilde{x}$

Algoritmo:

(i) ordenar os dados:

$$x_{(1)}, x_{(2)}, \cdots, x_{(n)}$$

(ii) calcular o termo central:

$$\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se } n \text{ é ímpar,} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se } n \text{ é par.} \end{cases}$$

Moda, $\hat{x}$

Se os dados não estão agrupados, o valor modal é o que ocorre mais vezes, se mais de um valor ocorre com igual frequência, ambos são modas.

Medidas de Dispersão

Variância, σ^2

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\overline{x})^2.$$

Desvio padrão, σ

$$\sigma = +\sqrt{\sigma^2}.$$

Coeficiente de variação, cov

$$cov = \frac{\sigma}{\overline{x}}.$$

Separatrizes

Seja $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ uma ordenação dos resultados em ordem crescente, conhecida como *estatśıstica de ordem* dos resultados. Então, em analogia com a definição da mediana, o quantil $q(1/n)$ é definido como sendo a média aritmética entre $x_{(1)}$ e $x_{(2)}$, de modo que exatamente $100/n\%$ dos resultados são menores que $q(1/n)$. Similarmente, o quantil $q(2/n)$ é definido como sendo a média aritmética entre $x_{(2)}$ e $x_{(3)}$. Mas, como $q(1/n) \leq x_{(2)} \leq q(2/n)$, o resultado $x_{(2)}$ deve corresponder a um quantil $q(p)$, onde $\frac{1}{n} < p < \frac{2}{n}$. Para a definição formal dos quantis assume-se linearidade entre os quantis da forma $q(m/n)$, para $m \leq n$. Então, como $x_{(2)} = \frac{q(1/n)+q(2/n)}{2}$, $x_{(2)}$ é igual ao quantil $q(\frac{\frac{1}{n}+\frac{2}{n}}{2}) = q(\frac{3}{2n})$. Em geral, seguindo o mesmo argumento, $x_{(i)}$ é igual ao quantil $q(\frac{\frac{i-1}{n}+\frac{i}{n}}{2}) = q(\frac{2i-1}{2n}) = q(\frac{i-0{,}5}{n})$, para $i = 1, 2, \ldots, n$.

Contudo, dependendo do valor de p, é preciso cuidado ao definir o quantil. Se $p < \frac{1}{2n}$, como $x_{(1)}$ é o menor valor observado dos resultados e é igual ao quantil $q(\frac{1}{2n})$, define-se $q(p)$ como sendo igual a $x_{(1)}$. Similarmente, se $p > \frac{2n-1}{2n}$, como $x_{(n)}$ é o maior valor observado dos resultados e é igual ao quantil $q(\frac{n-0{,}5}{n})$, define-se $q(p)$ como sendo igual a $x_{(n)}$. Finalmente, se $p = \alpha\frac{2(i-1)-1}{2n} + (1-\alpha)\frac{2i-1}{2n}$, onde $0 < \alpha < 1$, então define-se $q(p)$ como sendo igual a $\alpha x_{(i-1)} + (1-\alpha)x_{(i)}$.

Resumindo:

$$q(p) = \begin{cases} x_{(1)}, & \text{se } p < \frac{1}{2n}, \\ x_{(n)}, & \text{se } p > \frac{2n-1}{2n}, \\ x_{(i)}, & \text{se } p = \frac{2i-1}{2n}, \\ \alpha x_{(i-1)} + (1-\alpha)x_{(i)}, & \text{se } p = \alpha\frac{2(i-1)-1}{2n} + (1-\alpha)\frac{2i-1}{2n}, \text{ onde } 0 < \alpha < 1. \end{cases}$$

Exemplo 12.2.1: Considere que os resultados de um teste foram: 3,4,5,6, e 7. Determinar (a) $q(0.05)$, (b) $q(0.25)$, e (c) $q(0.75)$.

Solução: Para (a), como $0.05 < \frac{1}{10}$, tem-se que $q(0.05) = 3$. Para (b), note que $0.25 = \alpha(0.1)+(1-\alpha)0.3$, se $\alpha = 1/4$. Portanto, $q(0.25) = (1/4)3+(3/4)4 = 15/4$. Finalmente, para (c), note que $0.75 = \alpha(0.7) + (1-\alpha)0.9$, se $\alpha = 3/4$. Portanto, $q(0.75) = (3/4)6 + (1/4)7 = 25/4$. ■

Alternativamente, Jain (1991) propõe que a fórmula

$$(n-1)\alpha + 1$$

seja usada para calcular a posição da desejada separatriz, $q(p)$. Como é uma posição, então é um inteiro positivo, portanto, o valor encontrado deve ser arredondado para o inteiro mais próximo. Por exemplo, se $n = 356$ e deseja-se calcular C_{46}, isto é, o quadragésimo sexto centil, então a posição de $q(46)$ no conjunto de dados é dada por:

$$(356-1)\frac{46}{100} = 163.3$$

cujo inteiro mais próximo é 163. Portanto, $q(46)$ está na centésima sexuagésima terceira posição.

Uma das aplicações interessantes de separatrizes aparece em determinadas *Cartas de Recomendação* usualmente exigidas para ingresso de alunos em programas de pós-graduação. Geralmente nestas cartas é pedido que a pessoa que dará a referida carta de recomendação, o recomendante, classifique o candidato com respeito a sua aptidão para realizar estudos avançados e pesquisa dentre um total de n alunos dos últimos k anos. Assim, o recomendante deve escolher entre os 5% mais aptos ou 10% mais aptos ou 50% mais aptos, entre outros. Estes valores correspondem respectivamente a $q(0.95)$, $q(0.90)$ e $q(0.50)$.

Separatrizes também podem ser usadas para colocar conceitos de forma tão justa quanto possível. Os conceitos geralmente são A (excelente), B (bom), C (razoável), D (inaceitável). Se o professor usar a classificação $(9, 10]$, A, $(8, 9]$, B, $(7, 8]$, C e $[0, 7]$ D para uma prova (dificil!), onde nenhum aluno obteve nota maior que 8 então os conceitos estariam entre C e D. O professor pode usar um critério subjetivo para atribuir os conceitos que amenize a situação. Mas, o professor pode calcular separatrizes e colocar os conceitos justamente. Se $q(0.99)$ é computado, este valor corresponde ao limite inferior dos 1% melhores, portanto, independe de critérios subjetivos ou rótulos pré-estabelecidos. Então, uma classificação indenpendente do nível da prova poderia ser: $(q(0.99), 10]$, A; $(q(0.95), q(0.99)]$, B; $(q(0.80), q(0.95)]$, C; $[0, q(0.80)]$, D.

Exemplo 12.2.2: Analise os dados a seguir[1] que correspondem aos tempos de CPU de dado experimento. Utilize duas casas decimais e arredondamento simétrico[2].

3.1	4.2	2.8	5.1	2.8	4.4	5.6	3.9
3.9	2.7	4.1	3.6	3.1	4.5	3.8	2.9
3.4	3.3	2.8	4.5	4.9	5.3	1.9	4.2
3.2	4.1	5.1	3.2	3.9	4.8	5.9	4.2

Inicialmente os dados devem ser ordenados. A ordem usual é a crescente.

1.9	2.7	2.8	2.8	2.8	2.9	3.1	3.1
3.2	3.2	3.3	3.4	3.6	3.7	3.8	3.9
3.9	3.9	4.1	4.1	4.2	4.2	4.4	4.5
4.5	4.8	4.9	5.1	5.1	5.3	5.6	5.9

DADOS AGRUPADOS

Frequências para os tempos de CPU (ms)

Classes	frequências
$1.4 \vdash 2.4$	1
$2.4 \vdash 3.4$	10
$3.4 \vdash 4.4$	11
$4.4 \vdash 5.4$	8
$5.4 \vdash 6.4$	2
Total	32

[1] Jain (Jain, 1991), página 195.

[2] Isto é, o sistema de ponto-flutuante é $F(10, 2, e_1, e_2)$.

Utilizando as fórmulas vistas anteriormente, tem-se:

$$\overline{x} = \frac{124.8}{32} = 3.9,$$
$$\tilde{x} = 3.4 + \frac{16-11}{11} \times 1 = 3.85,$$
$$\hat{x} = 3.4 + \frac{1}{4} \times 1 = 3.65,$$
$$\sigma^2 = \frac{30}{32} \approx 0.94,$$
$$\sigma \approx 0.97,$$
$$q(0.25) = 2.4 + \frac{7}{10} \times 1 \approx 3.20,$$
$$q(\frac{3}{10}) = 2.4 + \frac{9.6-1}{10} \times 1 \approx 3.26,$$
$$q(0.90) = 4.4 + \frac{28.8-22}{8} \times 1 \approx 5.25.$$

Indicadores estatísticos para os tempos de CPU
(Cálculos com dados agrupados)

Indicador	Valor
Média aritmética	3.90ms
Mediana	3.85ms
Moda	3.65ms
Variância	0.94ms^2
Desvio padrão	0.97ms
1o. quartil	3.20ms
3o. decil	3.26ms
90o. percentil	5.25ms

DADOS NÃO-AGRUPADOS

O cálculo da média, moda, variância, desvio padrão e coeficiente de variação não necessitam que os dados estejam ordenados, mas as separatrizes, sim.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i =\approx 3.89,$$

$$\tilde{x} = \frac{3.9+3.8}{2} = 3.9,$$

$$\hat{x} = 2.8,\ \hat{x} = 3.9,$$

$$q(0.25) = \frac{1}{2}x_{(8)} + \frac{1}{2}x_{(9)} = 3.15 \approx 3.2,$$

$$q(0.9) = \frac{7}{10}x_{(29)} + \frac{3}{10}x_{(30)} = 5.16 \approx 5.2,$$

$$q(0.25) = x_{(9)} = 3.2,\ \text{(método Jain)}$$

$$q(0.90) = x_{(29)} = 5.1.\ \text{(método Jain)}$$

Indicadores estatísticos para os tempos de CPU
(Cálculos com dados não-agrupados)

Indicador	Valor
Média aritmética	3.89ms
Mediana	3.90ms
Moda	2.80ms e 3.90ms
1o. quartil	3.20ms ou 3.15ms
90o. percentil	5.10ms ou 5.16ms

■

Exemplo 12.2.3: O serviço de atendimento ao consumidor de uma concessionária de veículos recebe as reclamações dos clientes. Tendo em vista a melhoria na qualidade do atendimento foi anotado o número de reclamações diárias nos últimos 30 dias: 4, 5, 3, 4, 2, 6, 4, 1, 6, 5, 3, 4, 4, 5, 2, 3, 6, 5, 4, 2, 2, 3, 4, 3, 3, 2, 1, 1, 5, e 2.

(a) Faça uma tabela de frequências desses dados.

(b) Determine o valor da média, moda, mediana, desvio padrão, e do 1o. e 3o. quartis.

(c) Com base nos valores obtidos na letra (b), você diria que a distribuição dos dados é simé trica?

Solução: A tabela de frequências é dada por:

Tabela 12.5: Solução.

No. de Reclamações	Frequência relativa
1	3/30
2	6/30
3	6/30
4	7/30
5	5/30
6	3/30

A média dos dados é dada por:

$$\overline{x} = (1 \times 3/30) + (2 \times 6/30) + (3 \times 6/30) + (4 \times 7/30) + (5 \times 5/30) + (6 \times 3/30) \simeq 3.47.$$

A moda é igual a 4. A mediana é dada por 3.5. A variância é dada por:

$$\sigma^2 = (1\times3/30)+(4\times6/30)+(9\times6/30)+(16\times7/30)+(25\times5/30)+(36\times3/30)-3.47^2 \simeq 2.16.$$

Logo, o desvio padrão é igual aproximadamente a 1.47. O primeiro quartil é dado por $x_{(8)} = 2$, e o terceiro quartil é dado por $x_{(23)} = 5$. Com estes resultados observa-se que

(a) $q(0.5) - x_{(1)} = 2.5 = x_{(n)} - q(0.5)$;

(b) $q(0.5) - q(0.25) = 1.5 = q(0.75) - q(0.5)$;

(c) $q(0.25) - x_{(1)} = x_{(n)} - q(0.75)$.

Logo, estes dados formam uma distribuição simétrica. ■

Tabela 12.6: Dura ção de casamentos.

Anos de casamento	No. de Divórcios
$0 \vdash 6$	2800
$6 \vdash 12$	1400
$12 \vdash 18$	600
$18 \vdash 24$	150
$24 \vdash 30$	50

Exemplo 12.2.4: O número de divórcios na cidade, de acordo com a duração do casamento, está representado na Tabela 12.6:

(a) Encontre o 1o. e o 9o. decis.

(b) Qual o intervalo interquartil?

Solução:

(a)

$$q(0.1) = 0 + \frac{500 - 0}{2800} \times 6 = 1.07$$
$$q(0.9) = 12 + \frac{4500 - 4200}{600} \times 6 = 15$$

(b)

$$q(0.25) = 0 + \frac{1250 - 0}{2800} \times 6 = 2.68$$
$$q(0.75) = 6 + \frac{3750 - 2800}{1400} \times 6 = 10.07$$

Portanto, o intervalo interquartil é $[2.68, 10.07]$.

■

12.3 Exercícios

12.1 Gere k sequências de números aleatórios cada uma de tamanho n e as ordene, usando algoritmos de ordenação tais como o QuickSort, HeapSort e BubbleSort, entre outros. Analise os k resultados encontrados para cada algoritmo. Note que a variável de interesse é *tempo de execução* para cada algoritmo.

12.2 Analise o número de e-mails que chegam diaramente, semanalmente ou mensalmente a um servidor de determinada lista de discussão.

12.3 Foi construída uma distribuição de frequências com intervalos de classes de iguais amplitudes, sendo esta igual a 4, para 40 observações de peso de estudantes universitários, como na Tabela 12.7. Complete os valores que faltam.

Tabela 12.7: Distribuição de frequências para o peso de 40 estudantes universitários.

Classes	n_i	f_i	Fc_i	x_i
	8			
		0.175		
			26	53
	5			
			35	
		0.100		
Total			-	-

12.4 A Tabela 12.8 representa os salários pagos a 200 operários de uma dada empresa em número de salários mínimos. Complete-a e calcule o salário médio desses operários.

Tabela 12.8: Número de salários mínimos pagos a operários.

Classes	n_i	$Fc_i/200$	x_i
$\vdash$		0.15	1
$\vdash$		0.35	3
$\vdash$		0.45	5
$\vdash$		0.5	7
$\vdash$			9
Total			

Capítulo 13

Uma Introdução à Inferência Estatística

Neste capítulo serão abordados tópicos da Estatística, isto é, tópicos onde a suposição fundamental é suportada por métodos indutivos: uma amostra aleatória é retirada da população e, a partir desta, com um erro probabilístico fixado, asserções são realizadas sobre a população. Este é um método indutivo: do particular (amostra) induz-se para o geral (população). Na Probabilidade quando se resolve um problema sobre variáveis aleatórias, por exemplo, não se questiona qual o valor dos parâmetros de uma densidade, frases usuais são: seja uma Uniforme em (10,30), ou uma Exponencial de parâmetro α. O raciocínio probabilístico é abstrato e os métodos probabilísticos dedutivos. Entretanto, Probabilidade e Estatística se complementam para resolver problemas do mundo físco quando, por exemplo, o objetivo é descobrir qual a distribuição de um conjunto de observações, então métodos estatístico entram em cena para especificar, com determinado erro (probabilístico!), qual a melhor distribuição para o conjunto de observações.

Quando modelos probabilísticos são aplicados em algum problema prático, é preciso ter informação a respeito da distribuição de probabilidade da variável aleatória de interesse. Existem dois processos clássicos para a obtenção da distribuição de uma variável aleatória: eduzir uma distribuição a priori de um especialista da área, ou inferir a distribuição a partir de uma análise de dados. Neste livro, não serão tratados métodos de edução, e sim métodos de inferência.

13.1 População e Amostra

Suponha que o interesse de determinada pesquisa fosse a distribuição do consumo mensal de energia elétrica de todos os domicílios brasileiros. Se fosse possível obter os valores do consumo para todos os domicílios, a distribuição exata poderia ser obtida e daí calculados parâmetros de posição e dispersão, por exemplo. Nesse caso, inferência estatística não seria necessária pois seriam conhecidos todos os valores de interesse.

Porém, é raro a situação em que se consegue obter a distribuição exata de alguma variável, ou porque os custos são elevados, ou o tempo para a coleta de tais dados é longo, ou porque, às vezes, o experimento aleatório que se realiza consiste de um processo destrutivo. Por exemplo, medir a tensão máxima de entrada que um determinado tipo de estabilizador suporta. O experimento poderia começar com uma tensão de 0 volts, ir aumentado gradativamente até atingir a tensão máxima definida como sendo a tensão onde o estabilizador queimou. Deste modo, se todos os estabilizadores fossem testados, não restaria nenhum para ser vendido. Assim a solução é selecionar parte dos estabilizadores (amostra), analisá-la e inferir propriedades para todos os estabilizadores (população). Esta questão, dentre outras, é objeto de estudo da área de *inferência estatística*.

Definição 13.1.1: *População* é o conjunto de todos os elementos ou resultados de interesse.

Definição 13.1.2: *Amostra* é um subconjunto formado por elementos selecionados da população.

Frequentemente, usa-se um modelo probabilístico para uma população. Por exemplo, um cientista da computação pode considerar como normalmente distribuída, com média μ e variância σ^2 desconhecidas, a população de tempos de execução de determinado algoritmo; tem-se então uma população normal ou uma população distribuída normalmente. Como outro exemplo, suponha que o interesse seja investigar se uma dada moeda é honesta e para isso uma moeda é lançada 100 vezes. Neste caso, a população pode ser considerada como tendo a distribuição de uma variável aleatória X que assume o valor 1, se ocorrer cara, e 0 em caso contrário, que é uma Bernoulli com parâmetro p desconhecido. A amostra será a sequência binária de comprimento 100. Nestes dois últimos casos a população foi especificada como sendo a distribuição de uma variável aleatória X que modela a característica de interesse. Este método exige a proposta de um modelo para a variável X. São comuns as expressões "a população $f(x)$" ou "a população de tempos de execução $X \sim N(\mu, \sigma^2)$".

13.1.1 Seleção de uma Amostra

A fim de se ter inferências realmente informativas a respeito de uma dada população, precisa-se ter cuidado com os métodos de seleção de uma amostra; é necessário que a amostra seja *representativa* da população. Por exemplo, ao se fazer uma pesquisa de opinião pública a respeito de um dado governo, se só escolhidas pessoas que vivem em uma região beneficiada por esse governo, a amostra pode não ser representativa de toda a população, pois esta contém pessoas que foram beneficiadas pelo governo, neste caso diz-se que a amostra é *viesada*.

Nesta seção é visto apenas o caso de *amostragem aleatória simples*. Este procedimento é o método mais simples de se selecionar uma amostra aleatória de uma população e serve de base para outros métodos de amostragem mais complexos. No caso de uma população finita, implementa-se este método numerando os elementos da população e em seguida escolhendo um número numa tabela de números aleatórios ou gerando números aleatórios em um computador. Neste caso, todos os elementos da população têm a mesma probabilidade de serem selecionados. Repete-se o processo n vezes. A amostragem é *com reposição*, se um elemento da população puder ser escolhido mais de uma vez, e *sem reposição*, caso contrário. Apesar da amostragem sem reposição fornecer mais informação a respeito da população em estudo, a amostragem com reposição implica que cada elemento selecionado está sujeito ao mesmo mecanismo probabilístico e de modo independente uns dos outros, o que facilita a análise do modelo. Por isso este livro restringe-se ao caso com reposição. Em geral, tem-se a seguinte definição de amostra aleatória simples:

Definição 13.1.3: Dada uma população descrita por uma variável aleatória X, uma amostra aleatória simples de tamanho n é um conjunto de n variáveis aleatórias $(X_1, X_2, \ldots, X_n)$ independentes e identicamente distribuíidas cada uma com a mesma distribuição de X.

Intuitivamente, X_i representa a observação do i-ésimo elemento sorteado. Portanto, no caso de uma população X contínua, com função densidade de probabilidade f, a função densidade de probabilidade conjunta da amostra $(X_1, X_2, \ldots, X_n)$, será dada por:

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\cdots f(x_n).$$

Dependendo de como os números aleatórios são gerados, são conhecidos tanto a distribuição da variável aleatória sendo simulada quanto seus parâmetros. Por exemplo, ao se gerar 50 números de uma distribuição normal padrão obtém-se uma amostra aleatória simples de tamanho 50 desta população normal. Se outra pessoa observa apenas estes 50 números gerados, ela nada conhece a respeito da distribuição que os gerou nem seus parâmetros. O objetivo da inferência estatística é fornecer critérios para que se possa descobrir a forma da distribuição ou os parâmetros da população que gerou a amostra sendo observada.

13.2 Estatísticas e Parâmetros

Uma vez obtida a amostra de uma dada população, muitas vezes o interesse é calcular alguma função desta amostra. Por exemplo, a média da amostra $(X_1, X_2, \ldots, X_n)$ é dada por

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Como $\overline{X}$ é uma função de variáveis aleatórias, também é uma variável aleatória.

Definição 13.2.1: Uma *estatística* T é uma função da amostra $(X_1, X_2, \ldots, X_n)$.

As estatísticas mais comuns são:

(i) Média da amostra: $\overline{X} = (1/n)\sum_{i=1}^n X_i$;

(ii) Variância da amostra: $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 = \frac{1}{n-1}(\sum_{i=1}^n X_i^2 - n\overline{X}^2)$;

(iii) Proporção amostral: $\hat{p} = \frac{I_A(X_1)+\cdots+I_A(X_n)}{n}$, onde I_A é a função indicadora do evento de interesse A.

(iii) O menor valor da amostra: $X_{(1)} = \min(X_1, X_2, \ldots, X_n)$;

(iv) O maior valor da amostra: $X_{(n)} = \max(X_1, X_2, \ldots, X_n)$;

(v) Amplitude amostral: $W = X_{(n)} - X_{(1)}$;

(vi) A i-ésima maior observação da amostra: $X_{(i)}$.[1]

[1] Os elementos da amostra ordenados, isto é, $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$, são conhecidos como estatísticas de ordem da amostra.

Para diferenciar características da amostra de características da população, chama-se de *parâmetro* uma medida usada para descrever uma característica da população. Assim se uma população for modelada por uma variável aleatória X, a esperança e a variância $E(X)$ e $V(X)$, respectivamente, seriam parâmetros.

13.3 Distribuições Amostrais

Suponha o interesse em algum parâmetro θ da população e que decide-se usar uma estatística T de uma amostra aleatória simples $(X_1, X_2, \ldots, X_n)$ da população. Uma vez que a amostragem é realizada, pode-se calcular $T = t_0$ e é baseado neste valor que será realizada uma afirmação sobre θ. T sendo uma função de variáveis aleatórias também é uma variável aleatória e, portanto, possui uma dada distribuição. Esta distribuição é conhecida como *distribuição amostral da estatística T*.

Exemplo 13.3.1: Retiram-se com reposição todas as amostras de tamanho 2 da população $\{2, 3, 5, 7, 7\}$. Assim, o conjunto de todas as amostras de tamanho 2, com reposição, têm 25 elementos que são os pares $(2, 2)$, $(3, 5)$, entre outros. Estes pares vêm do produto cartesiano $\{2, 3, 5, 7, 7\} \times \{2, 3, 5, 7, 7\}$. A distribuição conjunta da amostra (X_1, X_2) é dada por:

	2	3	5	7
2	1/25	1/25	1/25	2/25
3	1/25	1/25	1/25	2/25
5	1/25	1/25	1/25	2/25
7	2/25	2/25	2/25	4/25

Calculando a distribuição da média amostral, $P(\overline{X} = 3) = P(X_1 = 3, X_2 = 3) = 1/25$. Similarmente, os demais valores podem ser obtidos conforme a Tabela 13.1:

Tabela 13.1: Distribuição da média amostral.

$\overline{x}$	2	2.5	3	3.5	4	4.5	5	6	7
$P(\overline{X} = \overline{x})$	1/25	2/25	1/25	2/25	2/25	4/25	5/25	4/25	4/25

■

Exemplo 13.3.2: Considere o lançamento de uma moeda 100 vezes, usando como estatística X o número de caras obtidas, a distribuição amostral desta estatística é uma binomial com parâmetros $n = 100$ e p, onde p é a probabilidade de cara em um lançamento qualquer desta moeda. Se o objetivo for saber se esta moeda é honesta, ou seja, se $p = 0.5$, e sabe-se que em 100 lançamentos ocorreram 64 caras, calcula-se, por uma $B(100, 0.5)$, $P_{0.5}(X \geq 64) = 0.0018$, ou seja, se a moeda for honesta, a probabilidade de se obterem 64 ou mais caras é igual a 0.0018, então existe evidência que p deve ser diferente de 0.5. Por outro lado, com 56 caras, obtém-se $P_{0.5}(X \geq 56) = 0.0867$, portanto, se a moeda for honesta aproximadamente $1/10$ das vezes observa-se um valor maior ou igual a 56, então, neste caso, os dados não são suficientes para descartar a hipótese que a moeda seja honesta. ■

Exemplo 13.3.3: Uma população consiste de quatro números 1, 3, 5, e 7. Considere todas as possíveis amostras de tamanho 2 de elementos que podem ser selecionadas com reposição desta população. Determine.

(a) A média e variância populacionais.

(b) A distribuição da média e variância amostrais.

Solução: A média populacional é dada por $\mu = \frac{1+3+5+7}{4} = 4$, e a variância populacional por $\sigma^2 = \frac{1^2+3^2+5^2+7^2}{4} - 4^2 = 5$. Para se determinar a média a variância amostrais, considere a Tabela 13.2 onde todas as possíveis amostras estão enumeradas:

Tabela 13.2: Todas as possíveis amostras.

x_1	x_2	$\overline{x}$	s^2
1	1	1	0
1	3	2	2
1	5	3	8
1	7	4	18
3	1	2	2
3	3	3	0
3	5	4	2
3	7	5	8
5	1	3	8
5	3	4	2
5	5	5	0
5	7	6	2
7	1	4	18
7	3	5	8
7	5	6	2
7	7	7	0

Como cada uma das possíveis amostras tem probabilidade $1/16$, a distribuição da média amostral e da variância amostral são respectivamente descritas pelas Tabelas 13.3 e 13.4:

e ■

Para algumas estatísticas não é possível obter analiticamente sua distribuição amostral, então simula-se um número grande de amostras diferentes e calculam-se as estatísticas de cada uma dessas amostras para obter uma distribuição amostral empírica da estatística de

Tabela 13.3: Distribuição da média Amostral

$\overline{x}$	1	2	3	4	5	6	7
$P(\overline{X} = \overline{x})$	1/16	2/16	3/16	4/16	3/16	2/16	1/16

Tabela 13.4: Distribuição da variância Amostral.

s^2	0	2	8	18
$P(S^2 = s^2)$	4/16	6/16	4/16	2/16

interesse. Por exemplo, para obter a mediana das temperaturas de amostras de 10 municípios retiradas da população $X \sim N(25, 25)$, pode-se gerar, via qualquer software, 1000 amostras de tamanho 10 desta população, determinar a mediana de cada uma dessas amostras e calcular medidas de posição e dispersão dos valores das medianas obtidos com essas amostras, bem como representação gráficas destes valores.

13.3.1 Distribuição da Média da Amostra, $\overline{X}$

A seguir será estudada a distribuição da média amostral $\overline{X}$. Antes de se obter informações sobre a forma desta distribuição, pode-se determinar a esperança e a variância de $\overline{X}$.

Teorema 13.3.4: *Seja X uma variável aleatória com média μ, variância σ^2 e $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória simples de X. Então,*

$$E(\overline{X}) = \mu \text{ e } V(\overline{X}) = \frac{\sigma^2}{n}.$$

Prova: Pela linearidade da esperança,

$$E(\overline{X}) = \frac{1}{n}(E(X_1) + E(X_2) + \cdots + E(X_n)) = \mu.$$

Como $X_1, X_2, \ldots, X_n$ são independentes,

$$V(\overline{X}) = \frac{1}{n^2}(V(X_1) + V(X_2) + \cdots + V(X_n)) = \frac{\sigma^2}{n}.$$

∎

Conforme n cresce a distribuição de $\overline{X}$ tende a ficar concentrada em torno de sua média μ, pois sua variância vai diminuindo. Além disso, o TCL, dá informação sobre a distribuição amostral da média para valores grandes de n.

Teorema 13.3.5: *Para amostras aleatórias simples* $(X_1, X_2, \ldots, X_n)$*, retiradas de uma população com média* μ *e variância* σ^2 *finita, a distribuição amostral da média* $\overline{X}$ *aproxima-se, para* n *grande, de uma distribuição normal, com média* μ *e variância* σ^2/n*, ou seja, se* $F_{\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}}$ *for a função de distribuição acumulada de* $\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}$*, então,* $\forall x \in I\!R$

$$\lim_n F_{\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}}(x) = \Phi(x).$$

Prova: A prova deste teorema está fora do escopo deste livro. Pode ser encontrada em B. James (1981). ▮

Caso a população já possua uma distribuição normal, como $\overline{X}$ é uma combinação linear de $X_1, X_2, \ldots, X_n$ que são independentes e possuem distribuição normal, a distribuição amostral da média amostral será exatamente uma normal para qualquer valor de n, e a média dessa distribuição será igual a média da população e variância será igual a variância da população dividida por n.

Em geral o TCL afirma que para valores grandes de n, $\overline{X}$ terá uma distribuição aproximadamente normal. Quão grande precisa ser n para que a distribuição de $\overline{X}$ seja aproximada por uma normal? Isto depende se a população tem uma distribuição próxima a normal ou não. Como regra empírica, para amostras de tamanho de 30 elementos, a aproximação para a normal já pode ser utilizada.

A diferença entre a média amostral e a média da população é conhecida como *erro amostral da média*, isto é, $e = \overline{X} - \mu$. O Teorema Central do Limite indica que $\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \sim N(0,1)$, ou seja, $\frac{\sqrt{n}e}{\sigma} \sim N(0,1)$.

Exemplo 13.3.6: Suponha que uma máquina está regulada para produzir lâmpadas com tempo de vida útil médio de 10000horas. De uma amostra de 50 lâmpadas produzidas por esta máquina verifica-se o tempo de vida útil de cada uma delas. Determine a probabilidade de que o tempo de vida útil médio seja menor ou igual a 8000horas.

Solução: Sabe-se que o tempo de vida útil de uma lâmpada é distribuído de acordo com uma exponencial. Portanto, como o tempo de vida útil médio é de 10000horas, a média populacional é 10000horas e a variância populacional é igual a 10^8horas^2. Além disso, como a amostra é maior que 30, utiliza-se o TCL para afirmar que a média amostral tem uma distribuição $N(10^4, \frac{10^8}{50})$. Portanto,

$$P(\overline{X} \leq 8000) = P(Z \leq \frac{\sqrt{50}(8000 - 10000)}{10000}) = \Phi(-\sqrt{2}) = 0.0793.$$

■

13.3.2 Distribuição da Variância da Amostra, S^2

No caso de uma população normal, o teorema a seguir fornece a distribuição da variância amostral.

Teorema 13.3.7: *Seja a variância amostral*

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Então,

$$E(S^2) = \sigma^2$$

e

$$(\frac{n-1}{\sigma^2})S^2 \sim \chi^2_{n-1},$$

desde que a população de onde foi retirada a amostra seja normalmente distribuída.

A prova deste teorema pode ser encontrada em Hogg e Craig (1970).

13.3.3 Distribuição da Proporção Amostral, $\hat{p}$

Suponha que a proporção de indivíduos de uma população que são portadores de uma determinada característica seja igual a p. Logo, pode-se definir uma variável aleatória X que é a função indicadora desta característica. Portanto, X tem uma distribuição Bernoulli de parâmetro p. Considere agora uma amostra aleatória simples de tamanho n desta população e seja S_n o número total de indivíduos na amostra que possuem a característica de interesse. Então, S_n tem uma distribuição binomial com parâmetros n e p. A proporção de indivíduos portadores da característica é dada por

$$\hat{p} = \frac{S_n}{n}.$$

Portanto, pode-se determinar a distribuição de $\hat{p}$ a partir da distribuição de S_n, utilizando a relação: $P(\hat{p} = \frac{k}{n}) = P(S_n = k)$.

Pelo Teorema Central do Limite se $(X_1, X_2, \ldots, X_n)$ formam uma amostra aleatória simples desta população, a distribuição amostral de $\overline{X}$ é aproximadamente igual a $N(p, p(1-p)/n)$ para valores grandes de n. Portanto, a distribuição de $S_n = n\overline{X}$ pode ser aproximada por uma normal $N(np, np(1-p))$. Como $\hat{p} = \overline{X}$, a distribuição da proporção amostral também pode ser aproximada por $N(p, p(1-p)/n)$ para valores grandes de n.

Exemplo 13.3.8: Uma máquina está regulada para produzir lâmpadas de modo que apenas 10% delas tenham tempo de vida útil menor ou igual a 1000horas. De uma amostra de 50 lâmpadas produzidas por esta máquina, qual a probabilidade de se encontrar no máximo 90% com tempo de vida útil maior que 1000 horas?
Solução: Como a amostra é maior que 30, utiliza-se o TCL para afirmar que a proporção amostral tem uma distribuição $N(0.1, \frac{(0.1)(0.9)}{50})$. Portanto,

$$P(1-\hat{p} \leq 0.9) = P(\hat{p} \geq 0.1) = P(Z \geq 0) = 0.5.$$

■

13.4 Estimadores e Estimativas

Um dos grandes interesses do estudo da Estatística é a obtenção de *estimativas* para os parâmetros populacionais. Se θ for um parâmetro desconhecido da população, deseja-se, a partir de uma função de uma amostra aleatória de tamanho n, $(X_1, X_2, \ldots, X_n)$, obter o valor mais plausível para o parâmetro θ. A estatística $\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$ utilizada para estimar θ é conhecida como *estimador*. Como $\hat{\Theta}$ é uma função da variável aleatória $\hat{\Theta}$ também é uma variável aleatória. Uma particular realização de $\hat{\Theta}$ é chamada de uma *estimativa* pontual para θ.

Os parâmetros populacionais mais comuns que se desejam estimar são:

(i) A média da população, μ.

(ii) A variância, σ^2, ou desvio-padrão σ, da população.

(iii) A proporção de itens populacionais que pertencem a uma classe de interesse, p.

(iv) A diferença de médias de duas populações, $\mu_1 - \mu_2$.

(v) A diferença de proporções de duas populações, $p_1 - p_2$.

Estimadores para esses parâmetros são, respectivamente:

(i) A média amostral, $\overline{X}$.

(ii) A variância amostral $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

(iii) A proporção amostral, $\hat{p}$.

(iv) A diferença de médias amostrais de duas amostras aleatórias independentes, $\overline{X}_1 - \overline{X}_2$.

(v) A diferença de proporções amostrais de duas amostras aleatórias independentes, $\hat{p}_1 - \hat{p}_2$.

Existem várias possibilidades de escolha para um estimador de um parâmetro. Por exemplo, o estimador $\frac{(n-1)S^2}{n}$ poderia ter sido escolhido para estimar a variância populacional. Portanto, é preciso estudar propriedades dos estimadores para desenvolver um critério que determine qual o melhor estimador para determinado parâmetro.

13.4.1 Propriedades de Estimadores

O problema da estimação é determinar um estimador que seja próximo de θ, segundo algum critério matemático. O primeiro critério é o seguinte:

Definição 13.4.1: O estimador T é *não-viesado* ou *não-tendencioso* para θ se $E(T) = \theta$.

O viés ou tendência de um estimador T para um parâmetro θ é igual a $E(T) - \theta$. Logo, um estimador T é não-viesado para θ, se o seu viés for igual a zero.

Exemplo 13.4.2: A média amostral $\overline{X}$ é um estimador não-viesado para média populacional μ, pois

$$E(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu.$$

■

Exemplo 13.4.3: A proporção amostral $\hat{p}$ é um estimador não-viesado para a proporção populacional p pois chamando de Y_i a variável aleatória que é igual a 1 se o i-ésimo indivíduo da amostra possui a característica de interesse, e igual a zero, em caso contrário, tem-se que

$$E(\hat{p}) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = p.$$

■

Exemplo 13.4.4: Considere uma população com N elementos, com média populacional $\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$, e variância populacional

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2.$$

Um possível estimador para σ^2, baseado numa amostra aleatória simples de tamanho n dessa população, é

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Entretanto, este estimador é viesado:

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \overline{X})^2 &= \sum_{i=1}^{n}(X_i - \mu + \mu - \overline{X})^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - 2\sum_{i=1}^{n}(X_i - \mu)(\overline{X} - \mu) + \sum_{i=1}^{n}(\overline{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\overline{X} - \mu)^2.
\end{aligned}$$

Portanto,

$$\begin{aligned}
E(\hat{\sigma}^2) &= \frac{1}{n}(\sum_{i=1}^{n} E(X_i - \mu)^2 - nE(\overline{X} - \mu)^2) \\
&= \frac{1}{n}(\sum_{i=1}^{n} V(X_i) - nVar(\overline{X})) \\
&= \frac{1}{n}(n\sigma^2 - n\frac{\sigma^2}{n}) = \frac{n-1}{n}\sigma^2.
\end{aligned}$$

Logo, o viés de $\hat{\sigma}^2$ é igual a $\frac{n-1}{n}\sigma^2 - \sigma^2 = \frac{-\sigma^2}{n}$. Portanto, o estimador $\hat{\sigma}^2$ em geral subestima o verdadeiro parâmetro σ^2. Por outro lado, o viés diminui quando n tende a infinito. É fácil ver que

$$S^2 = \frac{n}{n-1}\hat{\sigma}^2$$

é um estimador não-viesado para σ^2. Por isso, a variância de uma amostra de tamanho n é dada por S^2, onde o denominador é igual a $n-1$, enquanto que a variância de uma população de tamanho N é dada por σ^2, onde o denominador é igual a N. ∎

O segundo critério a ser analisado é o critério da consistência de um estimador. Intuitivamente, um estimador é consistente se quando aumenta-se o tamanho da amostra n, a probabilidade de que este difira do parâmetro por mais que qualquer erro pre-especificado $\epsilon > 0$ tende a zero. Formalmente,

Definição 13.4.5: Uma sequência $\{T_n\}$ de estimadores de um parâmetro θ é *consistente* se, para todo $\epsilon > 0$,

$$\lim_{n\to\infty} P(|T_n - \theta| > \epsilon) = 0.$$

Exemplo 13.4.6: A sequência de estimadores $\overline{X}_n$ é consistente, pois como $E(\overline{X}_n) = \mu$ e $V(\overline{X}_n) = \frac{\sigma^2}{n}$, utilizando a desigualdade de Tchebycheff:

$$P(|\overline{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \to 0,$$

quando $n \to \infty$, para qualquer $\epsilon > 0$. ∎

O seguinte teorema determina se uma dada sequência de estimadores é consistente:

Teorema 13.4.7: *Se $\{T_n\}$ é uma sequência de estimadores de θ tal que $\lim_{n\to\infty} E(T_n) = \theta$ e $\lim_{n\to\infty} V(T_n) = 0$, então $\{T_n\}$ é consistente.*

Prova: Pela desigualdade triangular, se $|T_n - \theta| > \epsilon$, então $|E(T_n) - \theta| > \frac{\epsilon}{2}$ ou $|T_n - E(T_n)| > \frac{\epsilon}{2}$. Portanto,

$$P(|T_n - \theta| > \epsilon) \leq P(|E(T_n) - \theta| > \frac{\epsilon}{2}) + P(|T_n - E(T_n)| > \frac{\epsilon}{2}).$$

Logo, pela desigualdade de Tchebycheff

$$P(|T_n - \theta| > \epsilon) \leq P(|E(T_n) - \theta| > \frac{\epsilon}{2}) + \frac{4V(T_n)}{\epsilon^2}.$$

Tomando os limites quando $n \to \infty$:

$$\lim_n P(|T_n - \theta| > \epsilon) \leq \lim_n P(|E(T_n) - \theta| > \frac{\epsilon}{2}) + \lim_n \frac{4V(T_n)}{\epsilon^2} = 0.$$

Portanto, $\{T_n\}$ é consistente. ∎

Se T_n for um estimador não-viesado, então obviamente $\lim_{n\to\infty} E(T_n) = \theta$, e portanto se a variância do estimador T_n tender a zero, ele é um estimador consistente.

Exemplo 13.4.8: Foi visto que S^2 é um estimador não-viesado para σ^2. É possível demonstrar no caso em que a população tem distribuição normal com média μ e variância σ^2 que

$$V(S^2) = \frac{2\sigma^4}{n-1}.$$

Logo, S^2 é consistente para σ^2. ■

Exemplo 13.4.9: Como $\hat{\sigma}^2 = \frac{n-1}{n}S^2$, então $E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 \to \sigma^2$ quando $n \to \infty$, e $V(\hat{\sigma}^2) = (\frac{n-1}{n})^2\frac{2\sigma^4}{n-1} \to 0$ quando $n \to \infty$. Logo, pelo teorema anterior, $\hat{\sigma}^2$ também é consistente para σ^2. ■

Um outro critério para comparação de estimadores é o seguinte:

Definição 13.4.10: Se T e T' são dois estimadores não-viesados de um mesmo parâmetro θ, e $V(T) < V(T')$, então T é mais *eficiente* que T'.

Exemplo 13.4.11: Considere uma população normal, X, com parâmetros μ e σ^2. O objetivo é estimar a mediana desta população. Como a distribuição é simétrica a mediana e a média coincidem e são iguais a μ. Definindo $\overline{X}$ e md, respectivamente, como a média e a mediana de uma amostra de tamanho n dessa população, qual dos dois estimadores é mais eficiente para estimar a mediana populacional?

Sabe-se que $\overline{X} \sim N(\mu, \sigma^2/n)$ e demonstra-se que a distribuição da mediana pode ser aproximada por $N(Md(X), \frac{\pi\sigma^2}{2n})$, onde Md é a mediana da população. Portanto, os dois estimadores são não-viesados, mas $\overline{X}$ é mais eficiente, pois $V(md) > V(\overline{X})$. Conclui-se que, para estimar a mediana dessa população, é preferível usar a média da amostra como estimador. ■

Finalmente, pode-se considerar o critério do *erro quadrático médio* para comparar estimadores.

Definição 13.4.12: Denomina-se *erro amostral* de um estimador T para um parâmetro θ a diferença $e = T - \theta$.

O erro amostral é uma variável aleatória pois é uma função de T que é variável aleatória; além disso, o viés de T é igual a esperança do erro amostral.

Definição 13.4.13: O *erro quadrático médio* (EQM) do estimador T para o parâmetro θ é igual ao segundo momento do erro amostral com respeito a distribuição amostral do estimador T, ou seja,

$$EQM(T, \theta) = E(e^2) = E(T - \theta)^2.$$

A expressão do EQM pode ser desenvolvida para:

$$
\begin{aligned}
&EQM(T,\theta) = E(T - E(T) + E(T) - \theta)^2 \\
&= E(T - E(T))^2 + 2E[(T - E(T))(E(T) - \theta)] + E(E(T) - \theta)^2 \\
&= V(T) + V^2.
\end{aligned}
$$

Então, o erro quadrático médio leva em consideração tanto o viés V do estimador como sua variabilidade medida através de $V(T)$. Segundo este critério, o estimador é tão melhor quanto menor for seu erro quadrático médio.

Exemplo 13.4.14: Determinar o erro quadrático médio do estimador $\overline{X}$ para μ.
Solução: Neste caso,

$$E(\overline{X} - \mu)^2 = V(\overline{X}) = \frac{\sigma^2}{n}.$$

■

13.4.2 Como obter Estimativas

Estimativas são instanciações dos estimadores. Os estimadores são funções, as estimativas são valores que as funções assumem.

Por exemplo, supondo uma população X com uma dada lei de probabilidade tal que $E(X) = \mu$. Foi visto anteriormente que um estimador para μ é $\overline{X}$, a média da amostra $(X_1, \ldots, X_n)$ retirada de X. Portanto, quando uma particular amostra é instanciada, tem-se um particular valor para $\overline{X}$, denominado de $\overline{x}$ e este é uma estimativa de $\overline{X}$. Da mesma forma, s^2 é uma estimativa de S^2.

13.5 Intervalos de Confiança

Os estimadores apresentados até o momento são pontuais, isto é, especificam uma única estimativa para o parâmetro. Para que se tenha uma medida de qual é a possível magnitude do erro que está sendo cometido é preciso de uma outra forma para obter estimativas. Uma alternativa é buscar um método para construir intervalos de números reais que contenham o parâmetro. Esses intervalos são os denominados *intervalos de confiança* e são baseados na distribuição amostral do estimador.

Um intervalo de confiança para um parâmetro populacional desconhecido θ é um intervalo da forma (L, U), em que os pontos extremos do intervalo L e U dependem da amostra, e portanto são, na verdade, estatísticas, isto é variáveis aleatórias. O objetivo ao se construir intervalos de confiança é determinar funções da amostra L e U tal que a seguinte afirmação seja verdadeira:

$$P(L \leq \theta \leq U) = 1 - \alpha,$$

onde $0 < \alpha < 1$. Assim, existe uma probabilidade $1 - \alpha$ de se selecionar uma amostra tal que o intervalo $[L, U]$ contenha o valor de θ. Note que θ não é aleatório, L e U é que são aleatórios. Se a afirmação acima for verdadeira o que está sendo dito é que se forem construídos vários intervalos de confiança usando as estimativas L e U, em $100(1-\alpha)\%$ das vezes θ estará incluso no intervalo $[L, U]$. Tal intervalo é chamado de um *intervalo de* $100(1-\alpha)\%$ *de confiança* para θ, e $(1-\alpha)$ é conhecido como *coeficiente de confiança* ou *nível de confiança* do intervalo e α é o *nível de significância*.

É importante destacar que é o método utilizado para determinar L e U que garante que com probabilidade $1-\alpha$, θ estará contido no intervalo $[L, U]$. Para uma particular amostra, o intervalo (l, u) observado pode ou não conter θ e não existe qualquer aleatoriedade uma vez que l, u e θ são números reais fixos. A afirmação correta é que (l, u) contém θ com $100(1-\alpha)\%$ de confiança.

Quanto maior o intervalo de confiança, mais confiança se tem que ele contenha o verdadeiro valor θ. Por outro lado, quanto maior for o intervalo, menos informação a respeito do verdadeiro valor de θ. O ideal seria ter um intervalo pequeno com alta confiança.

O intervalo de confiança descrito acima é um intervalo de confiança bilateral, pois são especificados tanto o limite inferior quanto o superior do intervalo. Pode-se também obter um intervalo de confiança unilateral inferior para θ com nível de confiança α, escolhendo um limite inferior L de tal forma que

$$P(L \leq \theta) = 1-\alpha.$$

Analogamente, um intervalo de confiança unilateral superior para θ com nível de confiança α, pode ser obtido escolhendo um limite superior U tal que

$$P(\theta \leq U) = 1-\alpha.$$

13.5.1 Intervalo de Confiança para a Média Populacional (μ) com Variância Populacional (σ^2) Conhecida

Pelo Teorema Central do Limite, a distribuição amostral de $\overline{X}$ é aproximadamente normal com média μ e variância σ^2/n, desde que n seja suficientemente grande ($n \geq 30$). Neste caso,

$$Z = \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}$$

tem uma distribuição normal padrão. Esta variável é facilmente usada para calcular um intervalo de confiança para μ.

Seja $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$. Tem-se que:

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha.$$

Logo,

$$P(-z_{\alpha/2} \leq \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \leq z_{\alpha/2}) = 1-\alpha.$$

Rearrumando as desigualdades com o objetivo de explicitar μ:

$$\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \leq z_{\alpha/2} \Rightarrow \mu \geq \overline{X} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}},$$

$$\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \geq -z_{\alpha/2} \Rightarrow \mu \leq \overline{X} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}}.$$

Desse modo, com $100(1-\alpha)\%$ de confiança,

$$(\overline{X} - \frac{z_{\alpha/2}\sigma}{\sqrt{n}}, \overline{X} + \frac{z_{\alpha/2}\sigma}{\sqrt{n}})$$

é um intervalo de confiança bilateral para μ.

Intervalos unilaterais podem ser construídos como a seguir. Fazendo $z_\alpha = \Phi^{-1}(1-\alpha)$:

$$P(Z \geq -z_\alpha) = 1-\alpha$$

e resolvendo a desigualdade abaixo para μ

$$\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma} \geq -z_\alpha \Rightarrow \mu \leq \overline{X} + \frac{z_\alpha\sigma}{\sqrt{n}}$$

obtem-se o intervalo unilateral *superior* com $100(1-\alpha)\%$ de confiança,

$$(-\infty, \overline{X} + \frac{z_\alpha\sigma}{\sqrt{n}}).$$

De forma similar,

$$P(Z \leq z_\alpha) = 1-\alpha \Rightarrow \mu \geq \overline{X} - \frac{z_\alpha\sigma}{\sqrt{n}},$$

e o intervalo unilateral *inferior* com $100(1-\alpha)\%$ de confiança é

$$(\overline{X} - \frac{z_\alpha\sigma}{\sqrt{n}}, +\infty).$$

Intervalos de confiança podem ser usados para determinar o tamanho de uma amostra que satisfaça determinada especificação.

13.5.2 Intervalo de Confiança para Média Populacional (μ) com Variância Populacional (σ^2) Desconhecida

Quando o objetivo é construir intervalos de confiança para a média μ de uma população quando σ^2 for desconhecida, devido ao Teorema Central do Limite, pode-se continuar usando os procedimentos da seção anterior, desde que o tamanho da amostra seja grande ($n \geq 30$), e que se adote s^2 como estimativa para σ^2.

Se a população tem uma distribuição normal, então $T = \frac{\sqrt{n}(\overline{X}-\mu)}{S}$ tem uma distribuição t de Student com $n-1$ graus de liberdade (Hogg e Craig, 1970). Portanto, fixado o nível de confiança $(1-\alpha)$ a partir do qual obtém-se o valor $t_{\alpha/2}$ da distribuição t de Student com

$n-1$ graus de liberdade e procedendo-se de forma similar ao que foi feito da seção anterior, tem-se que

$$P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = 1-\alpha$$

e portanto

$$(\overline{X} - \frac{t_{\alpha/2}S}{\sqrt{n}}, \overline{X} + \frac{t_{\alpha/2}S}{\sqrt{n}})$$

é um intervalo de confiança bilateral com $100(1-\alpha)\%$ de confiança para a média da população μ. Analogamente,

$$(-\infty, \overline{X} + \frac{t_{\alpha}S}{\sqrt{n}})$$

é um intervalo unilateral *superior* com $100(1-\alpha)\%$ de confiança para μ, e

$$(\overline{X} - \frac{t_{\alpha}S}{\sqrt{n}}, \infty)$$

é um intervalo unilateral inferior com $100(1-\alpha)\%$ de confiança para μ.

Exemplo 13.5.1: Seja uma população com distribuição Bernoulli de parâmetro p. Por exemplo, p pode representar a probabilidade de um tipo de capacitor ser produzido com defeito por uma determinada fábrica. Dada uma amostra aleatória $(X_1, X_2, \ldots, X_n)$ de tamanho n da produção de capacitores desta fábrica, pode-se estimar um intervalo de confiança bilateral para p. A variância da população é dada por $p(1-p)$. Portanto, sendo $\hat{p}$ a proporção de capacitores com defeito na amostra, como $\sigma^2 = p(1-p)$, então

$$(\hat{p} - \Phi^{-1}((\alpha+1)/2)\sqrt{\frac{p(1-p)}{n}}, \hat{p} + \Phi^{-1}((\alpha+1)/2)\sqrt{\frac{p(1-p)}{n}})$$

é um intervalo com $100\alpha\%$ de confiança para p. Como p não é conhecido, tem-se dois possíveis procedimentos para se obter seu valor:

(i) utilizar o fato de que $p(1-p) \leq 1/4$, obtendo o intervalo

$$(\hat{p} - \Phi^{-1}((\alpha+1)/2)\sqrt{\frac{1}{4n}}, \hat{p} + \Phi^{-1}((\alpha+1)/2)\sqrt{\frac{1}{4n}}),$$

(ii) usar $\hat{p}$ como estimativa para p, obtendo o intervalo

$$(\hat{p} - \Phi^{-1}((\alpha+1)/2)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + \Phi^{-1}((\alpha+1)/2)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}).$$

O primeiro método é sempre correto, porém muito conservador pois, em geral, $p(1-p)$ pode ser bem menor que $1/4$, e então o intervalo proposto tem amplitude maior que a necessária. O segundo método é válido desde que np e $n(1-p)$ sejam maiores que 5, pois, caso contrário, a distribuição normal não mais poderá ser usada sendo necessário utilizar a binomial. ∎

Exemplo 13.5.2: O comprimento dos eixos produzidos por uma empresa tem aproximadamente uma distribuição normal com desvio padrão $4cm$. Uma amostra com 16 eixos forneceu uma média de $4.52cm$. bach

(a) Determine um intervalo de confiança de 90% para o verdadeiro comprimento médio dos eixos.

(b) Com que probabilidade afirma-se que o comprimento médio desta amostra não difere da média por mais de $0.5cm$?

Solução: O intervalo de confiança é dado por:

$$(4.52 - \Phi^{-1}(0.95)\frac{4}{\sqrt{16}}, 4.52 + \Phi^{-1}(0.95)\frac{4}{\sqrt{16}}] = [2.875, 6.165).$$

Para o item (b), como $\sigma/\sqrt{n} = 1$, então $\overline{X} - \mu$ tem distribuição normal padrão, logo

$$P(|\overline{X} - \mu| \leq 0.5) = P(|Z| \leq 0.5) = 0.383.$$

■

Exemplo 13.5.3: Uma amostra de 400 domicílios mostra que 25% deles são alugados. Qual é o intervalo de confiança para o (verdadeiro) número de casas alugadas numa cidade, supondo que ela tem 20000 casas? Considere um coeficiente de confiança de 98%.
Solução: Inicialmente, determinando o intervalo de confiança para a proporção de casas alugadas. Neste caso, $\hat{p} = 0.25$, $n = 400$, e $\alpha = 0.98$. Utilizando $\hat{p}(1-\hat{p})$ como uma estimativa para a variância $p(1-p)$, o intervalo de confiança para a população é:

$$(0.25 - \Phi^{-1}(0.99)\sqrt{\frac{0.25(0.75)}{400}}, 0.25 + \Phi^{-1}(0.99)\sqrt{\frac{0.25(0.75)}{400}}).$$

Então, o intervalo de confiança para o número de casas alugadas é dado por:

$$(20000(0.25 - \Phi^{-1}(0.99)\sqrt{\frac{0.25(0.75)}{400}}), 20000(0{,}25 + \Phi^{-1}(0.99)\sqrt{\frac{0.25(0.75)}{400}}))$$

$$= (5000 - 1006.75, 5000 + 1006.75)$$
$$= (3993.25, 6006.75).$$

■

Exemplo 13.5.4: Uma pesquisa sobre renda familiar foi realizada entre as famílias que têm rendimento de até 5 salários mínimos. Sabe-se que o desvio padrão populacional é 1.2. Uma amostra de 200 famílias foi selecionada e seus resultados aparecem na tabela abaixo:

(a) Estime, com 95% de confiabilidade, o intervalo de confiança para a média de renda familiar desta população.

Tabela 13.5: Renda familiar.

Rendimento	Frequência
1	90
2	50
3	30
4	20
5	10

(b) Estime a verdadeira proporção de famílias que têm rendimento de até 2 salários mínimos, com 95% de confiabilidade.

Solução: Determinando inicialmente o valor de $\overline{x}$.

$$\overline{x} = 1(90/200) + 2(50/200) + 3(30/200) + 4(20/200) + 5(10/200) = 2.05.$$

Então, o intervalo de confiança de 95% é dado por:

$$(2.05 - \Phi^{-1}(0.975)\frac{1.2}{\sqrt{200}}, 2.05 + \Phi^{-1}(0.975)\frac{1.2}{\sqrt{200}}) = (1.884, 2.216).$$

Para o item (b), tem-se que $\hat{p} = 140/200 = 0.7$. Usando $\hat{p}(1-\hat{p})$ como estimativa para a variância populacional, o intervalo de confiança de 95% para proporção populacional é:

$$(0.7 - \Phi^{-1}(0.975)\sqrt{\frac{0.7(0.3)}{200}}, 0.7 + \Phi^{-1}(0.975)\sqrt{\frac{0.7(0.3)}{200}}) = (0.636, 0.764).$$

■

13.6 Teste de Hipóteses

Um teste de hipótese é um método para decidir se uma hipótese sobre um parâmetro populacional deve ou não ser aceita com base nos dados de uma amostra selecionada desta população. Ao contrário do que é estudado no problema de estimação, quando deseja-se estimar um parâmetro, como visto anteriormente, muitas vezes o objetivo é apenas checar se determinado parâmetro da população satisfaz uma condição de interesse. Tal condição é conhecida como *hipótese*. A idéia central deste procedimento é assumir que a hipótese é verdadeira e verificar se a amostra observada parece "razoável" ou "consistente", dada esta suposição.

Definição 13.6.1 : Uma *hipótese estatística* é uma condição que se quer testar sobre determinados parâmetros de uma ou mais populações.

No caso em que distribuições de probabilidade são usadas para representar populações, uma hipótese estatística também pode ser pensada como uma afirmação acerca destas.

Por exemplo, suponha que o interesse seja verificar a tensão em uma dada tomada. A tensão na tomada sofre alterações ao longo do dia e pode assim ser descrita por uma variável aleatória. Suponha que o interesse seja no valor esperado desta distribuição, ou seja, decidir se a tensão é ou não igual a $220v$. Então, $\mu = 220v$ é chamada de *hipótese nula*, representada por H_0. Esta hipótese nula pode ser aceita ou rejeitada; no caso de ser rejeitada, precisa-se de uma outra hipótese que seja aceitável, conhecida como *hipótese alternativa*, representada por H_1. Por exemplo, uma hipótese alternativa seria $\mu \neq 200v$. Neste caso, como a hipótese alternativa especifica valores de μ maiores ou menores que o valor especificado por H_0, ela é chamada de *hipótese alternativa bilateral.* Em algumas situações pode-se desejar formular uma *hipótese alternativa unilateral*, como em $H_0 : \mu = 220v$ e $H_1 : \mu < 220v$, $H_0 : \mu = 220v$ e $H_1 : \mu > 220v$, ou $H_0 : \mu = 220v$ e $H_1 : \mu = 240v$.

Então, a hipótese nula é uma afirmação a respeito da população, mais especificamente uma afirmação a respeito de um parâmetro da população. Esta afirmação pode ter sido originada de conhecimento *a priori* da população em estudo, de testes ou experimentos anteriores; pode ter sido determinada de alguma teoria ou modelo da população em estudo; ou pode surgir de considerações exógenas, por exemplo, parâmetros que devem obedecer certos critérios de controle de qualidade.

Estabelecidas as hipóteses nulas e alternativas, a informação contida na amostra é analisada para verificar se a hipótese nula é consistente com esta informação. Caso seja, conclui-se que a hipótese nula é verdadeira, caso contrário, conclui-se que a hipótese é falsa, o que implicará na aceitação da hipótese alternativa. Porém, para se saber com certeza se a hipótese nula é ou não verdadeira, seria necessário analisar toda a população, o que na prática é frequententemente impossível. Portanto, todo procedimento de testes de hipóteses tem de ter alguma probabilidade de erro associada, uma vez que é fundamentado numa amostra de tamanho n.

Para ilustrar, considere o exemplo descrito anteriormente, ou seja, $H_0 : \mu = 220v$ e $H_1 : \mu = 240v$. Suponha que n medidas na tensão da tomada sejam feitas e que a média dos valores obtidos nesta amostra $\overline{x}$ seja observada. Como visto, $\overline{x}$ é uma estimativa para o valor de μ, logo se for obtido um valor de $\overline{x}$ próximo a $220v$, tem-se uma evidência de que a hipótese nula é verdadeira. Precisa-se então estabelecer uma região de valores, conhecida como *região de aceitação* tal que se $\overline{x}$ cair nesta região a hipótese nula é aceita, e se $\overline{x}$ cair fora dessa região, ou seja, na região conhecida como *região crítica* (RC), a hipótese alternativa é aceita. Por exemplo, pode-se considerar a região de aceitação como sendo o intervalo $(-\infty, 230]$. Os limites da região de aceitação são chamados de *valores críticos.*

Esse procedimento de decisão acarreta um de dois tipos de erros diferentes. O primeiro, conhecido como *erro tipo I* ocorre quando a tensão média na tomada é realmente $220v$, mas por chance o conjunto de medidas aleatórios obtido forneceu um valor de $\overline{x}$ na região crítica. Ou seja, um erro do tipo I ocorre quando a hipótese nula é rejeitada quando na verdade ela é verdadeira. O segundo, conhecido como *erro do tipo II* ocorre quando apesar da hipótese nula ser falsa, a média das medidas de tensão obtidas cai na região de aceitação. Ou seja, um erro do tipo II ocorre sempre que a hipótese nula for aceita mesmo sendo falsa.

A probabilidade de ocorrência de um erro tipo I é chamada de *nível de significância*, *tamanho do teste*, ou ainda, *p-valor do teste*, e é denotada por α. O *poder de um teste* é igual a probabilidade de se rejeitar a hipótese nula quando ela realmente é falsa. Note que o poder do teste é igual a 1 menos a probabilidade de ocorrência de um erro do tipo II, que é usualmente denotada por β.

Quando H_0 for verdadeira, isto é, a tensão for realmente de $220v$, pelo TCL sabe-se que $\overline{X} \sim N(220, \frac{\sigma^2}{n})$. Então, o nível de significância do teste é determinado por:

$$\begin{aligned}\alpha = P(\text{erro I}) &= P(\overline{X} > 230|\overline{X} \sim N(220, \frac{\sigma^2}{n}))\\ &= P(\frac{\sqrt{n}(\overline{X}-220)}{\sigma} > \frac{\sqrt{n}(230-220)}{\sigma})\end{aligned}$$

Se a variância da tensão na tomada é $64v^2$, com uma amostra de 4 medidas do valor de tensão obtém-se:

$$\alpha = P(Z > \frac{2(10)}{8}) = P(Z > 2.5) = 0.0062.$$

De modo análogo, obtém-se a probabilidade do erro tipo II. Se H_1 for verdadeira, $\overline{X} \sim N(240, 16)$, então:

$$\begin{aligned}\beta = P(\text{erro II}) &= P(\overline{X} \leq 230|\overline{X} \sim N(240, 16))\\ &= P(\frac{(\overline{X}-240)}{4} \leq \frac{(230-240)}{4}) = P(Z \leq -2.5) = 0.0062.\end{aligned}$$

Neste caso, α e β foram iguais devido a simetria da região crítica em relação às hipóteses nula e alternativa. Se ao invés do valor crítico ser 230, fosse maior, então α diminuiria e β aumentaria.

Também seria possível especificar um valor para a probabilidade de erro do tipo I e verificar qual seria a região crítica que satisfaria esta probabilidade de erro pre-especificada. Por exemplo, suponha que se queira encontrar a região crítica cujo α seja igual a 0.01. Então:

$$0.01 = \alpha = P(Z > 2.325) = P(\frac{2(\overline{X}-220)}{8} > 2.325) = P(\overline{X} > 229.3).$$

Para a região crítica $(229.3, \infty)$, o valor de β é:

$$\begin{aligned}\beta = P(\text{erro II}) &= P(\overline{X} \leq 229.3|\overline{X} \sim N(240, 16))\\ &= P(\frac{(\overline{X}-240)}{4} \leq \frac{(229.3-240)}{4}) = P(Z \leq -2.675) = 0.0038.\end{aligned}$$

Este segundo tipo de procedimento é bastante utilizado, pois em geral a hipótese alternativa não contém apenas um único valor de parâmetro como no exemplo acima. Muitas vezes, se a hipótese nula é $H_0 : \mu = 220v$, a hipótese alternativa será $H_1 : \mu \neq 220v$. Como os parâmetros da hipótese alternativa são muitos, a solução é adotar o último procedimento descrito acima, ou seja, pre-estabelecer um valor α, e calcular uma região crítica que satisfaça esta restrição. No caso de uma hipótese alternativa bilateral, em geral toma-se como região de aceitação um intervalo simétrico ao redor da hipótese nula, deste modo fixando $\alpha = 0.01$,

$$0.01 = \alpha = P(|Z| > 2.575) = P(|\frac{2(\overline{X} - 220)}{8}| > 2.575) = 1 - P(209.7 \leq \overline{X} \leq 230.3).$$

Portanto, a região de aceitação $(209.7, 230.3)$ foi determinada de modo que o nível de significância de 0.01 fosse satisfeito. Mesmo determinada esta regra de decisão, não determina-se β, pois não existe um único valor de μ na hipótese alternativa. Neste caso, considera-se uma função $\beta(\mu)$, conhecida como *função característica de operação.*

Definição 13.6.2: A *função característica de operação* (função CO) de um teste de hipótese é definida por:

$$\beta(\mu) = P(\text{aceitar } H_0|\mu).$$

Isto é, $\beta(\mu)$ é a probabilidade de aceitar H_0 como função de μ.

Definição 13.6.3: A *função poder do teste*, é dada por $\pi(\mu) = 1 - \beta(\mu)$.

Portanto, esta função é a probabilidade de se rejeitar H_0 como função de μ. As seguintes propriedades de $\pi(\mu)$ são facilmente verificadas:

(i) $\pi(\mu_0) = \alpha$;

(ii) No caso de hipótese alternativa bilateral ($H_1 : \mu \neq \mu_0$), $\pi(-\infty) = \pi(+\infty) = 1$ e $\pi(\mu)$ decresce para $\mu < \mu_0$ e cresce para $\mu > \mu_0$;

(iii) No caso de hipótese alternativa unilateral superior ($H_1 : \mu > \mu_0$), $\pi(-\infty) = 0$, $\pi(+\infty) = 1$, e $\pi(\mu)$ é sempre crescente;

(iv) No caso de hipótese alternativa unilateral inferior ($H_1 : \mu < \mu_0$), $\pi(-\infty) = 1$, $\pi(+\infty) = 0$, e $\pi(\mu)$ é sempre decrescente.

Na definição das hipóteses, sempre estabelece-se a hipótese nula como uma igualdade, de modo que o analista pode controlar α, ao estabelecer uma região crítica para o teste. Assim, o analista pode controlar diretamente a probabilidade de rejeitar erroneamente H_0, implicando que a rejeição da hipótese nula é uma *conclusão forte*. Note que quanto menor o valor de α, quando a hipótese nula é rejeitada, mais provável é a hipótese alternativa, portanto maior será a significância da conclusão. Por isso, α é chamado de nível de significância do teste. Por outro lado, β não é constante, mas depende do verdadeiro valor do parâmetro, por este motivo a aceitação de H_0 é tida como uma *conclusão fraca*, a não ser que saiba-se que β é aceitavelmente pequena. Então, a nomenclatura mais correta seria ao invés de se dizer H_0

é aceita deveria ser dito *a amostra não apresentou evidência suficiente para se rejeitar* H_0. Neste último caso, não necessariamente afirma-se que existe uma alta probabilidade de que H_0 seja verdadeira, isto pode significar apenas que mais dados são necessários para que uma conclusão tomada esteja mais próxima da realidade.

Na determinação de quem é a hipótese nula, deve-se adotar como H_0 aquela hipótese, que se rejeitada erroneamente, conduza a um erro mais importante de se evitar, pois esta probabilidade de erro é controlável. Então, por exemplo, se o interesse é saber se um novo medicamento é eficaz no combate a uma doença, a hipótese nula seria que ele é não eficaz, pois os danos causados por ser usado um remédio não eficaz são maiores que se um remédio que seria eficaz não fosse usado. Ou ainda, se se deseja saber se certa substância é radioativa, então a hipótese nula seria que ela é radioativa, pois os danos causados pela manipulação radioativa são maiores que se uma substância não fossse manipulada por se achar falsamente que ela é radioativa. Como a rejeição da hipótese nula é que é uma conclusão forte, escolhe-se como H_1 a hipótese que se deseja comprovar. Por exemplo, no caso do novo medicamento H_1 será a hipótese que o novo medicamento é melhor que os existentes.

13.6.1 Procedimento para realizar um Teste de Hipótese

Segue-se uma sequência de passos para a realização de qualquer teste de hipótese:

(i) Identifique o parâmetro de interesse no problema em análise.

(ii) Fixe qual a hipótese nula H_0 e alternativa H_1.

(iii) Use teoria estatística e informações disponíveis para decidir que estimador será usado para testar H_0.

(iv) Obtenha a distribuição do estimador proposto.

(v) Determine α.

(vi) Construa a região crítica para o teste de modo que α seja satisfeita.

(vii) Use os dados da amostra para determinar o valor do estimador, ou seja, uma estimativa para o parâmetro.

(viii) Se o valor do estimador pertencer a região crítica, rejeite H_0. Caso contrário, reporte que não existe evidência suficiente para se rejeitar H_0.

13.6.2 Teste de Hipótese para a Média de uma População Normal com Variância Conhecida

Deseja-se testar as hipóteses $H_0 : \mu = \mu_0$ e $H_1 : \mu \neq \mu_0$, sendo μ_0 uma constante especificada. Para testar a hipótese nula, usa-se o estimador média amostral de uma amostra aleatória simples de tamanho n. Deste modo, se a hipótese nula for verdadeira, pelo TCL, $\overline{X} \sim N(\mu_0, \sigma^2/n)$ e então procede-se como anteriormente.

A estatística padronizada $Z_0 = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ tem uma distribuição normal padrão, se a hipótese nula for verdadeira. Portanto, para a região de aceitação

$$(-\Phi^{-1}(1-\alpha/2), \Phi^{-1}(1-\alpha/2)),$$

tem-se que $P(Z_0 \in RC|\mu = \mu_0) = \alpha$.

É mais fácil entender a região crítica e o procedimento do teste quando a estatística de teste é Z_0 e não $\overline{X}$. Entretanto, a mesma região crítica pode ser calculada em termos do valor da estatística $\overline{X}$. Neste caso, a região de aceitação é

$$(\mu_0 - \Phi^{-1}(1-\alpha/2)\frac{\sigma}{\sqrt{n}}, \mu_0 + \Phi^{-1}(1-\alpha/2)\frac{\sigma}{\sqrt{n}}).$$

De modo similar, obtém-se a região crítica para o caso de um teste de hipótese unilateral $H_0 : \mu = \mu_0$ e $H_1 : \mu > \mu_0$, ou $H_0 : \mu = \mu_0$ e $H_1 : \mu < \mu_0$. No primeiro caso, a região de aceitação para a estatística Z_0 é

$$(-\infty, \Phi^{-1}(1-\alpha)),$$

o que implica que a região de aceitação para a estatística $\overline{X}$ é

$$(-\infty, \mu_0 + \Phi^{-1}(1-\alpha)\frac{\sigma}{\sqrt{n}}).$$

No segundo caso, a região para a estatística Z_0 é

$$(\Phi^{-1}(\alpha), \infty),$$

o que implica que a região de aceitação para a estatística $\overline{X}$ é

$$(\mu_0 + \Phi^{-1}(\alpha)\frac{\sigma}{\sqrt{n}}, \infty).$$

13.6.3 Teste para a Proporção

O teste para proporção é um caso particular do caso do teste para a média com variância conhecida. Cada amostra pode ser considerada como uma variável Bernoulli com parâmetro p que representa a proporção de indivíduos da população que possuem uma determinada característica. Sabe-se que a média de uma Bernoulli é igual ao seu parâmetro p e que sua variância é igual a $p(1-p)$. Logo, utilizando a proporção amostral como estatística e os resultados gerais da seção anterior, a região de aceitação para a proporção é

(i) No caso de hipótese alternativa bilateral: $H_0 : p = p_0$ e $H_1 : p \neq p_0$, a região de aceitação é

$$(p_0 - \Phi^{-1}(1-\alpha/2)\sqrt{\frac{p_0(1-p_0)}{n}}, p_0 + \Phi^{-1}(1-\alpha/2)\sqrt{\frac{p_0(1-p_0)}{n}}).$$

(ii) No caso de hipótese alternativa unilateral superior: $H_0 : p = p_0$ e $H_1 : p > p_0$, a região de aceitação é

$$(-\infty, p_0 + \Phi^{-1}(1-\alpha)\sqrt{\frac{p_0(1-p_0)}{n}}).$$

(iii) No caso de hipótese alternativa unilateral inferior: $H_0 : p = p_0$ e $H_1 : p < p_0$, a região de aceitação é

$$(p_0 + \Phi^{-1}(\alpha)\sqrt{\frac{p_0(1-p_0)}{n}}, \infty).$$

Exemplo 13.6.4: Um relatório afirma que 40% de toda água obtida através de poços artesianos é salobra. Existem controvérsias sobre esta afirmação, alguns dizem que a proporção é maior outros que é menor. Para acabar com a dúvida, sorteou-se 400 poços e observou-se que em 120 deles a água era salobra. Qual deve ser a conclusão ao nível de significância de 3%?
Solução: Neste caso, $H_0 : p = 0.4$ contra uma hipótese alternativa bilateral $H_1 : p \neq 0.4$. Logo, a região de aceitação é dada por:

$$\begin{aligned}(0.4 - \Phi^{-1}(0.985)\sqrt{\frac{(0.4)(0.6)}{400}}, 0.4 + \Phi^{-1}(0.985)\sqrt{\frac{(0.4)(0.6)}{400}}) &= (0.4 - 0.053, 0{,}4 + 0.053) \\ &= (0.347, 0.453).\end{aligned}$$

Como $\hat{p} = 120/400 = 0.3$, rejeita-se a hipótese nula ao nível de confiança de 3%. ■

Exemplo 13.6.5: O governo afirma que a taxa de desemprego da população economicamente ativa é de no máximo 15% Uma amostra aleatória de 1500 pessoas revelou que 1300 destas estão empregadas. Para um nível de significância de 5%, pode-se dizer que a afirmação está correta?
Solução: Neste caso, a hipótese nula é $H_0 : p = 0.15$ contra a alternativa $H_1 : p < 0.15$. Logo, a região de aceitação é dada por:

$$(0.15 + \Phi^{-1}(0.05)\sqrt{\frac{(0.15)(0.85)}{1500}}, \infty) = (0.135, +\infty).$$

Como $\hat{p} = 200/1500 = 0.133$, rejeita-se a hipótese nula ao nível de confiança de 5%, e portanto a afirmação estava correta. ■

13.6.4 Testes para Amostras Grandes

Quando $n \geq 30$ a variância da amostra s^2 é próxima de σ^2, assim s pode ser usado no lugar de σ nos procedimentos anteriores. Deste modo, o teste para a média de uma população com variância conhecida pode ser utilizado, no caso de $n \geq 30$, para testar a média de uma população com variância desconhecida. O tratamento exato no caso em que σ^2 é desconhecida e a amostra é pequena envolve o uso da distribuição t de Student, quando a população é normalmente distribuída e será estudado a seguir.

13.6.5 Teste para a Média de uma População Normal com Variância Desconhecida

Assim como no caso de intervalos de confiança, quando a amostra for pequena e σ^2 desconhecida, uma suposição sobre a forma da distribuição em estudo tem de ser feita. Assume-se que a população tem uma distribuição normal e portanto $T = \frac{\sqrt{n}(\overline{X}-\mu)}{S}$ tem uma distribuição t de Student com $n-1$ graus de liberdade. Seja $t_{\alpha,n-1}$ o valor tal que $P(T \leq t_{\alpha,n-1}) = 1-\alpha$. Então, utilizando um procedimento similar ao caso de variância conhecida, se a estatística utilizada for a média amostral $\overline{X}$,

(i) No caso de hipótese alternativa bilateral: $H_0 : \mu = \mu_0$ e $H_1 : \mu \neq \mu_0$, a região de aceitação é

$$(\mu_0 - (t_{\alpha/2,n-1})\frac{S}{\sqrt{n}}, \mu_0 + (t_{\alpha/2,n-1})\frac{S}{\sqrt{n}}).$$

(ii) No caso de hipótese alternativa unilateral superior: $H_0 : \mu = \mu_0$ e $H_1 : \mu > \mu_0$, a região de aceitação é

$$(-\infty, \mu_0 + (t_{\alpha/2,n-1})\frac{S}{\sqrt{n}}).$$

(iii) No caso de hipótese alternativa unilateral inferior: $H_0 : \mu = \mu_0$ e $H_1 : \mu < \mu_0$, a região de aceitação é

$$(\mu_0 + (t_{\alpha/2,n-1})\frac{S}{\sqrt{n}}, \infty).$$

Exemplo 13.6.6: Uma rede de *fast food* pretende instalar uma nova lanchonete em certo local se nele transitarem no mínimo 200 carros por hora durante determinados períodos do dia. Para 20 horas selecionadas aleatoriamente durante tais períodos, o número médio de carros que transitou pelo lugar foi de 208.5 com desvio padrão de 30.0. O gerente assume a hipótese de que o volume de carro não satisfaz a exigência de 200 ou mais carros por hora. Para um nível de significância de 5% esta hipótese pode ser rejeitada?
Solução: A hipótese nula é dada por $H_0 : \mu = 200$ e a alternativa, $H_1 : \mu > 200$. Como a amostra é pequena ($n < 30$)) e a variância da população é desconhecida, deve-se usar o teste t de Student unilateral superior. Assim, a região de aceitação é dada por:

$$(-\infty, 200 + t_{0.95,19}\frac{30}{\sqrt{20}}) = (-\infty, 200 + 1.729\frac{30}{\sqrt{20}}) = (-\infty, 211.6).$$

Portanto, a hipótese não pode ser rejeitada a este nível de confiança. ∎

Exemplo 13.6.7: Num estudo sobre a resistência de um dado material, com distribuição normal, foi coletada uma amostra de 25 unidades, resultando num valor médio de 230.4kg e desvio-padrão de 100kg. O estudo é para saber se essa amostra é suficiente para garantir ao nível de significância de 5% que a resistência média do material seja superior a 200kg. Qual a sua conclusão?

Solução: O estudo quer realizar o seguinte teste: $H_0 : \mu = 200$ contra $H_1 : \mu > 200$. Como a variância é desconhecida e a amostra é menor que 30, usa-se o teste t de Student. A região de aceitação é

$$(-\infty, 200 + t_{0.95,24}\frac{100}{\sqrt{25}}) = (-\infty, 234.2).$$

Logo, a amostra não é grande o suficiente para se garantir que a resistência média é maior que 200 ao nível de significância de 5%. ■

13.6.6 Probabilidade de Significância

O procedimento do testes de hipóteses descrito até agora parte de pre-estabelecimento de um valor para α. Deste modo, como a escolha de α é arbitrária pode acontecer que para um determinado valor de α a hipótese nula seja rejeitada, porém para um valor menor de α ela não seja rejeitada. Além disso, no procedimento descrito, se a estimativa do parâmetro caía na região crítica a hipótese nula era rejeitada sem nenhuma informação a respeito de quão próxima essa estimativa estava da região de aceitação. Uma maneira alternativa para que tais problemas sejam evitados consiste em apresentar a *probabilidade de significância*, *nível descritivo*, ou *p-valor* do teste. Os passos são muito parecidos, só que ao invés de se construir a região crítica, apresenta-se o valor da probabilidade de ocorrerem valores da estatística mais extremos que o observado quando a hipótese nula é verdadeira. O p-valor também pode ser definido como o menor nível de significância que conduz a rejeição da hipótese nula com os dados observados.

Suponha que o interesse seja um teste para a média de uma população com variância conhecida (ou então variância desconhecida, mas amostra grande). Seja $\overline{x}_0$ a média amostral observada na amostra. Para um teste bilateral $H_0 : \mu = \mu_0$ e $H_1 : \mu \neq \mu_0$, tem-se

$$p = P(|\overline{X} - \mu_0| > |\overline{x}_0 - \mu_0|) = P(\frac{\sqrt{n}|\overline{X} - \mu_0|}{\sigma} > \frac{\sqrt{n}|\overline{x}_0 - \mu_0|}{\sigma})$$
$$= P(|Z| > \frac{\sqrt{n}|\overline{x}_0 - \mu_0|}{\sigma}) = 2(1 - \Phi(\frac{\sqrt{n}|\overline{x}_0 - \mu_0|}{\sigma})).$$

Similarmente, para um teste unilateral superior $H_0 : \mu = \mu_0$ e $H_1 : \mu > \mu_0$:

$$p = P(\overline{X} > \overline{x}_0) = P(\frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} > \frac{\sqrt{n}(\overline{x}_0 - \mu_0)}{\sigma})$$
$$= P(Z > \frac{\sqrt{n}(\overline{x}_0 - \mu_0)}{\sigma}) = 1 - \Phi(\frac{\sqrt{n}(\overline{x}_0 - \mu_0)}{\sigma}).$$

Tabela 13.6: Escala de evidência de Fisher.

p-valor	0.1	0.05	0.025	0.01	0.005	0.001
Natureza da Evidência	marginal	moderada	substancial	forte	muito forte	fortíssima

Finalmente, para um teste unilateral inferior $H_0 : \mu = \mu_0$ e $H_1 : \mu < \mu_0$:

$$p = P(\overline{X} < \overline{x}_0) = P(\frac{\sqrt{n}(\overline{X} - \mu_0)}{\sigma} < \frac{\sqrt{n}(\overline{x}_0 - \mu_0)}{\sigma})$$
$$= P(Z < \frac{\sqrt{n}(\overline{x}_0 - \mu_0)}{\sigma}) = \Phi(\frac{\sqrt{n}|\overline{x}_0 - \mu_0|}{\sigma}).$$

Exemplo 13.6.8: Suponha novamente a situação anterior onde deseja-se testar a hipótese nula $H_0 : \mu = 220v$ *versus* $H_1 : \mu \neq 220v$, com uma amostra de tamanho 4 e sabe-se que a variância é igual a $64v^2$. Suponha ainda que a média amostral foi igual a $227v$. O p-valor pode ser calculado por:

$$p = 2(1 - \Phi(\frac{\sqrt{4}|227 - 220|}{\sqrt{64}})) = 2(1 - \Phi(\frac{7}{4})) = 2(1 - 0.9599) = 0.0802.$$

Portanto, a probabilidade de quando a hipótese nula é verdadeira uma amostra selecionada de tamanho 4 tenha média amostral mais distante de $220v$ que $227v$ é igual a 0.0802, ou ainda, a um nível de significância de 10% a hipótese nula seria rejeitada, mas a um nível de significância de 5% a hipótese nula não pode ser rejeitada. ∎

A hipótese H_0 será rejeitada se o p-valor for "bastante pequeno". A Tabela 13.6 ilustra a escala de evidências de Fisher contra a hipótese H_0:

13.6.7 Significância Estatística *versus* Significância Prática

Quando o procedimento de um teste de hipótese é aplicado na prática, é necessário, além de se considerar a significância estatística medida pelo p-valor, analisar quais diferenças entre valores dos parâmetros têm implicações práticas. Isto é, pode acontecer que o p-valor seja pequeno levando então a rejeição da hipótese H_0, mas que o desvio real entre o valor do parâmetro na hipótese nula e a estimativa do parâmetro obtida na amostra não tenha *significância prática*. Isto pode ocorrer para tamanhos de amostras grandes. Por exemplo, para uma amostra de 1600 medidas e média amostral de $220.5v$ o p-valor bilateral é

$$p = 2(1 - \Phi(\frac{\sqrt{1600}(|220.5 - 220|)}{\sqrt{64}})) = 2(1 - \Phi(20/8)) = 0.0124.$$

Portanto, existe uma evidência estatística substancial para se rejeitar H_0. Contudo, do ponto de vista prático, se a média for realmente for $220.5v$ não haverá efeito prático observável no desempenho de qualquer equipamento elétrico. Logo, esta diferença detectada pelo teste de hipótese apesar de ter significância estatística não tem significância prática.

É preciso ter cuidado ao se interpretar os resultados de um teste de hipótese principalmente quando a amostra tiver tamanho grande, pois qualquer desvio pequeno do valor do parâmetro testado na hipótese nula será detectado como tendo significância estatística pelo teste, contudo, como visto, esta diferença poderá ter pouca ou nenhuma significância prática.

13.7 Teste de Aderência ou Bondade de Ajuste

Um objetivo comum em aplicações da Estatística em problemas da vida real é conhecer qual a distribuição de probabilidade de um conjunto de dados[2]. Este problema é resolvido aplicando um teste de hipótese proposto por K. Pearson.

Testar a aderência entre dados amostrais e populacionais resume-se em testar se valores amostrais e populacionais deferem significativamente estre si. Seja $A_1, \ldots, A_k$ uma partição do espaço amostral de um experimento aleatório tal que $P(A_i) = p_i$. Suponha que este experimento é repetido n vezes e que f_{o_i} é o número de amostras dentro do evento A_i. Definindo $f_{e_i} = np_i$ como o número esperado de amostras dentro de A_i, a estatística do teste

$$Q^2 = \sum_{i=1}^{k} \frac{(f_{o_i} - f_{e_i})^2}{f_{e_i}}$$

tem uma distribuição que pode ser aproximada por uma χ^2, com $k-1$ graus de liberdade, se não for necessário estimar os parâmetros populacionais, ou $k-1-r$, se os r parâmetros populacionais tiverem de ser estimados. Por exemplo, se a suposição for que a população segue uma Poisson, $r = 1$, se Normal, $r = 2$.

Portanto,

$$H_0 = f_{o_i} = f_{e_i}, \forall i,$$

e o teste consiste em calcular Q^2 e comparar com o valor de uma χ^2 com $k-1$ ou $k-1-r$ graus de liberdade e $100(1-\alpha)\%$ nível de confiança.

O numerador de Q^2 envolve a diferença de quadrados entre as frequências observadas e esperadas. As frequências teóricas vêm da população, são valores teóricos, não podem ser nulos. Também, se forem muitos pequenos, Q^2 apenas refletirá a mgnitude de f_{o_i}. Portanto, a literatura sugere que as frequências esperadas sejam pelo menos 3.

Exemplo 13.7.1: O número de defeitos por frequência, (x_i, f_{o_i}), numa amostra de 60 brinquedos de determinado fabricante se distribui como abaixo. O objetivo é testar se os dados podem seguir uma Poisson.

x_i	f_{o_i}
0	32
1	15
2	9
3	4

[2]Na verdade, o conjunto de dados são os valores observados de uma, ou mais, variável aleatória, portanto, a distribuição de probabilidade relevante é a da variável aleatória que gerou o conjunto de dados.

1. *Resolvendo o problema sem especificar completamente a distribuição de probabilidade.* Isto significa que supõe-se uma distribuição de probabilidade para os dados mas *não se conhece seus parâmetros.* Os passos para a realização do teste são:

   (i) Variável do problema, X, número de defeitos nos brinquedos analisados.

   (ii) H_0: os dados seguem uma Poisson,
   H_1: os dados não seguem uma Poisson.

   (iii) A estimativa para o parâmetro da Poisson é a média amostral (uma vez que a média da Poisson é também seu parâmetro). Logo, $\overline{x} = ((0 \times 32) + \cdots + (3 \times 4))/60 = 0.75$ e portanto

   $$X \sim \frac{e^{-0.75} 0.75^k}{k!}, k = 0, 1, \ldots$$

   (iv) Calculando as probabilidades teóricas:

   $$p_0 = P(X = 0) = 0.47,$$

   $$p_1 = P(X = 1) = 0.35,$$

   $$p_2 = P(X = 2) = 0.13,$$

   $$p_3 = P(X = 3) = 0.13,$$

   $$\ldots$$

   (v) Calculando as frequências esperadas, $f_{e_i} = np_i$, $i = 0, \ldots, 3$ e $n = 60$,

   $$f_{e_0} = 28.2, f_{e_1} = 21, f_{e_2} = 7.8, f_{e_3} = 2.4.$$

   Como $f_{e_3} < 3$, as duas últimas frequências esperadas são adicionadas. Portanto, para o problema, $f_{e_0} = 28.2$, $f_{e_1} = 21$, $f_{e_2} = 10.2$.

   (vi) Calculando a estatística do teste, Q^2.

   $$Q^2 = \sum_{i=0}^{2} \frac{(f_{o_i} - f_{e_i})^2}{f_{e_i}} = 2.99.$$

   (vii) Comparando Q^2 com o valor de uma χ^2 com $k - 1 - r = 3 - 1 - 1 = 1$ graus de liberdade e com nível de confiança de 95%.

   $$\chi^2_{1,95\%} = 3.841,$$

   $$Q^2 = 2.99$$

   $$Q^2 < \chi^2_{1,95\%} \Rightarrow$$

   H_0 não pode ser rejeitada.

2. *Resolvendo o problema especificando completamente a distribuição de probabilidade.* Neste caso, o parâmetro da variável aleatória populacional é conhecido. No problema em questão a suposição é que os dados seguem uma Poisson de parâmetro 0.75. Na realização do teste é o valor (tabelado) da χ^2 com $k-1=3-1=2$ graus de liberdade e 95% de confiança é 5.991, o que fornece a mesma conclusão anterior.

■

Se a variável for contínua, para o cálculo das frequências esperadas, a reta é dividida em intervalos mutuamente exclusivos e, a seguir, calculadas as probabilidades teóricas.

13.8 Aprendendo um pouco mais

Os resultados abaixo sumarizam os intervalos de confiança bilaterais mais usados em problemas práticos. Os intervalos em (iii), (iv), (v), (vi) e (vii) são obtidos de forma similar a que foi usada em (i) e (ii). O fundamento para a construção dos intervalos de confiança a seguir é que tem-se uma população, representada pela variável aleatória X, a qual se distribui normalmente com média μ_X e variância σ_X^2, isto é, $X \sim N(\mu_X, \sigma_X^2)$, e desta população foi retirada uma amostra aleatória $(X_1, X_2, \ldots, X_n)$. Enfatiza-se que a distribuição de probabilidade do estimador tem de ser conhecida. Por exemplo, se o objetivo é encontrar uma estimativa para a média da população, utiliza-se, obviamente, a média da amostra e a distribuição da média da amostra tem de ser conhecida.

Se a população não é normal mas a amostra é grande pelo TCL a distribuição normal pode ser usada e portanto os resultados abaixo ainda valem.

(i) Para a Média (μ_X), com Variância (σ_X^2) conhecida

$$(\overline{X} - z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}), \text{ onde}$$

$$P(N(0,1) \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}, \text{ e}$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

(ii) Para a Média (μ_X), com Variância (σ_X^2) desconhecida

$$(\overline{X} - t_{\alpha/2}\frac{S_X}{\sqrt{n}}, \overline{X} + t_{\alpha/2}\frac{S_X}{\sqrt{n}}), \text{ onde}$$

$$P(t_{n-1} \leq t_{\alpha/2}) = 1 - \frac{\alpha}{2}, \text{ e}$$

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

(iii) Para a Variância (σ_X^2)

$$\frac{(n-1)S_X^2}{\sigma_{n-1}^2} \sim \chi_{n-1}^2,$$

$$P(a < \chi_{n-1}^2 < b) = 1 - \alpha \Rightarrow \sigma^2 \in (\frac{(n-1)S_X^2}{b}, \frac{(n-1)S_X^2}{a}),$$

$$P(\chi_{n-1}^2 \leq a) = \frac{\alpha}{2}.$$

$$P(\chi_{n-1}^2 \leq b) = 1 - \frac{\alpha}{2}.$$

(iv) Para a Diferença de Médias ($\mu_X - \mu_Y$), com Variâncias (σ_X^2, σ_Y^2) conhecidas

$$\overline{X} - \overline{Y} \sim N(\mu_X - \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{n}).$$

$$((\overline{X} - \overline{Y}) - z_{\alpha/2}\sqrt{\frac{m\sigma_X^2}{n\sigma_Y^2}}, (\overline{X} - \overline{Y}) + z_{\alpha/2}\sqrt{\frac{m\sigma_X^2}{n\sigma_Y^2}}).$$

$$P(N(0,1) \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}.$$

(v) Para o Quociente de Variâncias ($\frac{\sigma_X^2}{\sigma_Y^2}$)

$$X \sim N(\mu_X, \sigma_X^2);\ Y \sim N(\mu_Y, \sigma_Y^2)$$

$$\mu_X, \sigma_X^2, \mu_Y, \sigma_Y^2 \text{ desconhecidos}$$

$$(X_1, \ldots, X_n); (Y_1, \ldots, Y_m)$$

$$\frac{(n-1)S_X^2}{\sigma_X^2} \sim \chi_{n-1}^2.$$

$$\frac{(m-1)S_Y^2}{\sigma_Y^2} \sim \chi_{m-1}^2.$$

Logo,

$$\frac{S_X^2}{\sigma_X^2}\frac{S_Y^2}{\sigma_Y^2} \sim F_{n-1,m-1}.$$

$$P(a < F_{n-1,m-1} < b) = 1 - \alpha \Rightarrow \frac{\sigma_X^2}{\sigma_Y^2} \in (\frac{S_X^2}{bS_Y^2}; \frac{S_X^2}{aS_Y^2}),$$

$$a = F_{n-1,m-1,\frac{\alpha}{2}};\ b = F_{n-1,m-1,1-\frac{\alpha}{2}}.$$

(vi) Para a Diferença de Médias com Variâncias desconhecidas, mas iguais

$$S^2 = \frac{S_X^2}{n} + \frac{S_Y^2}{m} \Rightarrow S = +\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}.$$

$$(\overline{X} - t_{\alpha/2}S, \overline{X} + t_{\alpha/2}S),$$

$$P(t_{\alpha/2,(n+m-2)} \leq t) = \frac{\alpha}{2},$$

onde $n + m - 2$ é o número de graus de liberdade.

(vii) Para a Proporção (p)

$$\hat{p} = \frac{X_n}{n}.$$

$$(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}).$$

$$P(N(0,1) \leq z_{\alpha/2)} = \frac{\alpha}{2}.$$

13.9 Exercícios

13.1 Rubens Barrichelo, nos treinos para a temporada de 2000, pilotando uma Ferrari no circuito de Magny Cours na França, fez uma média de um minuto e cincoenta segundos por volta, com um desvio padrão de 45 segundos. Sabendo-se que uma corrida nesse circuito tem 66 voltas, calcule:

(a) Qual a probabilidade dessa corrida ultrapassar o tempo limite de 2 horas?

(b) Qual o tempo provável da melhor e da pior volta, com 90% de confiança de acerto? (proposto por Dalton César P. Shibuya)

13.2 Os tempos de execução de um determinado algoritmo obtidos através da replicação de um experimento de simulação foram: 1.5, 2.0, 3.4, 1.8, 2.5 e 5.0. Com 80% de confiança em que região se encontra seu tempo médio de execução?

13.3 Um analista de sistema precisa decidir sobre a eficiência de um programa desenvolvido recentemente. Resolve adotar a seguinte regra de decisão: executar o programa 6 vezes para conjuntos de dados escolhidos aleatoriamente e construir um intervalo de confiança de 98% para o tempo médio de execução do programa; considerar o programa eficiente se a amplitude do intervalo de confiança obtido for menor do que 4.5 ms. Qual foi a decisão do administrador se os tempos amostrais, em milisegundos (ms), obtidos foram: 228, 230, 232, 229, 231, 230?

13.4 Os tempos de execução, em segundos, de 40 programas processados em um centro de processamento de dados foram

10	19	90	40	15	11	32	17	4	152
23	13	36	101	2	14	2	43	32	15
27	1	57	17	3	30	50	4	62	48
9	11	20	13	38	54	46	12	5	26

Encontre um intervalo de confiança a um nível de significância de 10% para o verdadeiro tempo médio de execução dos programas.

13.5 O gerente de um CPD sabe que o número de linhas de código, X, de um programa tem uma distribuição normal com variância 81. Recentemente este CPD contratou 36 novos estagiários e o gerente resolveu mandá-los, cada um independentemente do outro, otimizar o número de linhas de código do programa. O número médio de linhas decorrente do trabalho dos estagiários foi 100. Encontre um intervalo de confiança de 90% para o número médio de linhas (na população).

13.6 Um pesquisador está estudando a resistência de um determinado material. Ele sabe que essa variável é normalmente distribuída com um desvio padrão de 2 unidades.

(a) Utilizando os valores a seguir obtidos de uma amostra, 4.9, 7.0, 8.1, 4.5, 5.6, 6.8, 7.2, 6.2 determine o intervalo de confiança para a resistência média, com um nível de confiança de 0.90.

(b) Qual o tamanho da amostra necessário se quisermos que o erro cometido, ao estimar a resistência média, não seja superior a 0.01 unidades com probabilidade 0.90?

13.7 $\hat{\theta}$ é um estimador não-tendencioso para um parâmetro populacional θ se $E(\hat{\theta}) = \theta$. Seja uma amostra aleatória $(X_1, \cdots, X_n)$ de uma população cuja distribuição de probabilidade tem esperança e variância, respectivamente, μ e σ^2. Quais dos estimadores abaixo são não-tendenciosos para μ?

(a) $\hat{\theta}_1 = \frac{X_1+X_n}{2}$.

(b) $\hat{\theta}_2 = \frac{X_1+X_n}{n}$.

(c) $\hat{\theta}_3 = \frac{X_1+\cdots+X_n}{n}$.

(d) Qual a distribuição de $\hat{\theta}_3$, se a distribuição de probabilidade da população é normal e n é grande?

13.8 Em um exaustivo teste de vida para 10 componentes que não são vendidas com garantia, os tempos de falha observados, em horas, foram 1200, 1500, 1625, 1725, 1750, 1785, 1800, 1865, 1900 e 1950. Encontre um intervalo de confiança de 90% para o tempo médio populacional.

13.9 As diferenças no tempo de processamento entre duas diferentes implementações de um mesmo algoritmo foram mensuradas sobre 7 *workloads* resultando 1.5, 2.6, -1.8, 1.3, -0.5, 1.7 e 2.4.

(a) Você pode afirmar, com 90% de confiança que uma implementação é superior a outra?

(b) Teste se a diferença entre as medidas é igual a 1 com 99% de confiança.

13.10 Seis similares *workloads* foram usadas em dois sistemas. As observações foram (5.4,19.1), (16.6, 3.5), (0.6,3.4), (1.4,2.5), (0.6,3.6), (7.3,1.7). Você consideraria um sistema melhor que outro com 95% de confiança?

13.11 Um experimento consistiu de medir 32 vezes o tempo de uso da CPU por um determinado *software*. Os resultados foram 3.1, 4.2, 2.8, 5.1, 2.8, 4.4, 5.6, 3.9, 3.9, 2.7, 4.1, 3.6, 3.1, 4.5, 3.8, 2.9, 3.4, 3.3, 2.8, 4.5, 4.9, 5.3, 1.9, 3.7, 3.2, 4.1, 5.1, 3.2, 3.9, 4.8, 5.9, 4.2. Encontre um intervalo de confiança para a média com 90% de confiança.

13.12 As diferenças entre os valores medidos e os valores preditos usando um modelo analítico para um sistema é chamada de *modelagem do erro*. A *modelagem do erro* para um dado sistema forneceu os seguintes valores -0.04, -0.19, 0.14, -0.09, -0.14, 0.19, 0.04, 0.09. Encontre um intervalo de confiança para a média com 95% de confiança.

13.13 O tempo requerido para executar determinada tarefa foi medido em dois sistemas A e B. Os tempos para o sistema A foram 5.36,16.57, 0.62, 1.41, 0.64, 7.26; para o sistema B, 19.12, 3.52, 3.38, 2.50, 3.60, 1.74. Ao nível de 90% você consideraria os dois sistemas estatisticamente distintos?

13.14 A mesma *workload* foi aplicada 40 vezes a dois sistemas A e B. Constatou-se que o sistema A foi superior ao B 26 vezes. Será que podemos afirmar com 99% de confiança que o sistema A é superior?

13.15 Uma população tem desvio padrão igual a 10.

(a) Que tamanho deveria ter uma amostra para que, com probabilidade 8%, o erro em estimar a média seja inferior a 1?

(b) Supondo-se colhida a amostra no caso anterior, qual o intervalo de confiança para a média populacional, se a média amostral é 50?

13.16 A Tabela 13.7 contém os desvios com respeito aos diâmetros de vários cilindros produzidos por uma determinada máquina. Teste a hipótese de que as observações obedecem à lei normal se o nível de significância de 5% é usado.

Tabela 13.7: Diâmetro de cilindros.

limites dos intervalos em microns	0-5	5-10	10-15	15-20	20-25
n_i	15	75	100	50	20
p_i	0.06	0.30	0.40	0.20	0.04

13.17 Uma substância radioativa é observada durante 2608 iguais intervalos de tempo, cada um com 7.5 segundos. Para cada um dos intervalos de tempo, foi anotado o número de partículas detectador por um contador. Os números m_i de intervalos de tempo durante os quais i partículas alcançaram o contador são dados na tabela abaixo:

Tabela 13.8: Tempo de chegada de partículas ao contador.

i	0	1	2	3	4	5	6	7	8	9	≥ 10
m_i	57	203	383	525	532	408	273	139	45	27	16

Teste, usando um teste qui-quadrado, a hipótese de que os dados concordam com uma lei de Poisson. O nível de significância deve ser tomado como sendo 5%.

13.18 Na Tabela 13.9 estão listados m_i lotes de igual área $(0.25km^2)$, na parte Sul de Londres, cada um dos quais, durante a II Guerra Mundial foi acertado por i bombas. Teste, com a ajuda da distribuição qui-quadrado que os dados concordam com uma distribuição de Poisson, se o nível de significância de 6% é usado.

13.19 Para uma fina camada de solução com ouro, anotou-se o número de partículas de ouro que alcançaram o campo de visão do microscópio, durante iguais intervalos de tempo, conforme a Tabela 13.10. Teste, através de um teste-de-bondade-de-ajuste qui-quadrado, usando 5% de significância, que os dados seguem uma lei de Poisson.

Tabela 13.9: Número de bombas em lotes de Londres na II Guerra Mundial.

i	0	1	2	3	4	≥ 5
m_i	229	211	93	35	7	1

Tabela 13.10: Número de partículas de ouro que alcançaram o microscópio.

número de partículas, i	0	1	2	3	4	5	6	7
m_i	112	168	130	68	32	5	1	1

13.20 Dez tiros foram disparados por um rifle em cem alvos. O número de acertos foi registrado, na TabelařefTabelaTiro. Teste se as probabilidades de acerto nos alvos foram as mesmas em todos os tiros; em outras palavras, teste se os dados obedecem a uma distribuição binomial. Use um nível de significância de 10%.

Tabela 13.11: Número de acertos de tiros.

número de acertos, i	0	1	2	3	4	5	6	7	8	9	10
m_i	0	2	4	10	22	26	18	12	4	2	0

13.21 Sete moedas foram lançadas simultaneamente 1536 vêzes, e, cada vez, o número de caras, X, foi registrado. Os dados constam na Tabela 13.12. Usando um teste qui-quadrado e um nível de significância de 5% teste se os dados experimentais obedecem a uma distribuição binomial. Assuma que a probabilidade de ocorrência de cara é 0.5 em cada moeda.

13.22 Suponha que 250 números foram gerados somando-se os 5 dígitos de 250 números escolhidos de uma tabela de números aleatórios. Os resultados foram divididos em 15 intervalos e são mostrados na Tabela 13.13. Use um teste qui-quadrado para testar se a distribuição estatística dos dados segue uma distribuição normal. Use um nível de significância de 5%.

13.23 Os dígitos $0, 1, 2, \cdots 9$ entre os primeiros 800 casas decimais do número φ ocorrem 74, 92, 83, 79, 80, 73, 77, 75 e 91 vêzes, respectivamente. Use um teste qui-quadrado para testar a hipótese de que estes dados obedecem a uma lei uniforme. Considere o nível de significância de 10%.

13.24 De uma tabela de números aleatórios, 150 números de dois dígitos foram selecionados (00 também é um número de dois dígitos). Os resultados aparecem na Tabela 13.14.

Tabela 13.12: Número de caras.

X_i	0	1	2	3	4	5	6	7
m_i	12	78	270	456	385	252	69	13

Tabela 13.13: Soma de dígitos.

intervalo	m_i	intervalo	m_i	intervalo	m_i
0-3	0	15-18	28.5	30-33	27.0
3-6	0.5	18-21	39.0	33-36	7.5
6-9	1.5	21-24	41.0	36-39	1.0
9-12	10.0	24-27	45.0	39-42	1.0
12-15	17.5	27-30	30.5	42-45	0

Use um teste qui-quadrado para testar a hiótese de que estes dados obedecem a uma lei uniforme, com um nível de significância de 5%.

Tabela 13.14: Números de dois dígitos.

intervalo	m_i	freq. relativa	intervalo	m_i	freq. relativa
0-9	16	0.107	50-59	19	0.127
10-19	15	0.100	60-69	14	0.093
20-29	19	0.127	70-79	11	0.073
30-39	13	0.087	80-89	13	0.087
40-49	14	0.093	90-99	16	0.107

Referências Bibliográficas

1. **R. J. Adler, R. E. Feldman, and M. S. Taqqu** (1998). *A Practical Guide to Heavy Tails. Statistical Techniques and Applications*. Birkhäuser.

2. **A. O. Allen** (1978). *Probability, Statistics, and Queueing Theory with Computer Science Applications*. Academic Press.

3. **G. Ávila** (1994). *Cálculo–Funções de Várias Variáveis*. Volume 2. Livros Técnicos e Científicos Editora.

4. **L. Breiman** (1969). *Probability and Stochastic Processes*. Houghton Mifflin Company.

5. **C. W. Burrill** (1972). *Measure, Integration, and Probability*. McGraw-Hill.

6. **M. A. Campos** (1997). *Uma Extensão Intervalar para a Probabilidade Real*. Tese. Pós-Graduação em Ciência da Computação, Centro de Informática/UFPE.

7. **K. L. Chung** (1974). *A Course in Probability Theory*. Academic Press, Second Edition.

8. **K. L. Chung** (1974). *Elementary Probability Theory with Stochastic Processes*. Springer-Verlag.

9. **A. B. Clarke e R. L. Disney** (1979). *Probabilidade e Processos Estocásticos*. Livros Técnicos e Científicos.

10. **D. M. Cláudio e J. M. Marins** (1994) *Cálculo Numérico Computacional. Teoria e Prática*. Editora Atlas.

11. **W. J. Cody** (1988). *Floating-Point Standards, Theory and Practice*. Reliability in Computing. The Role of Interval Methods in Scientific Computing, Editor R. E. Moore, pages 99-108, Vol. 19.

12. **P. L. de O. Costa Neto e M. Cymbalista** (1974). *Probabilidades: resumos teóricos, exercícios resolvidos, exercícios propostos*. Edgard Blücher.

13. **D. Ellsberg** (1961). *Risk, Anbiguity, and the Savage Axioms*. Quaterly Journal of Economics 75 (4): 643-669. DOI: 10.2307/1884324.

14. **W. Feller** (1976). *Introdução à Teoria das Probabilidades e suas Aplicações*. Edgard Blücher Ltda.

15. **W. Feller** (1967). *An Introduction to Probability Theory and Its Applications*. Volume I. Third Edition. John Wiley & Sons.

16. **W. Feller** (1971). *An Introduction to Probability Theory and Its Applications*. Volume II. Second Edition. John Wiley & Sons.

17. **P. J. Fernandez** (1973). *Introdução à Teoria das Probabilidades*. Livros Técnicos e Científicos.

18. **T. Fine** (2005). *Probability and Probabilistic Reasoning for Electrical Engineers*. Prentice Hall.

19. **G. E. Forsythe** (1977). *Pitfalls in Computation, or Why a Math Book Isn't Enough.* Amer. Math. Monthly, pp. 931-955.

20. **D. Goldberg** (1991). *What Every Computer Scientist Should Know About Floating-Point Arithmetic.* ACM Computing Surveys, Vol. 23, No. 1, pp 5-48.

21. **B. V. Gnedenko** (1969). *The Theory of Probability.* Mir Publishers.

22. **C. M. Grinstead and J. L. Snell** (1991). *Introduction to Probability.* American Mathematical Society.

23. **H. L. Guidorizzi** (1994). *Um Curso de Cálculo.* Volume 4. Livros Técnicos e Científicos Editora.

24. **J. Halpern** (2003). *Reasoning about Uncertainty.* The MIT Press.

25. **P. G. Hoel, S. C. Port and C. J. Stone** (1978). *Introdução à Teoria da Probabilidade.* Interciência.

26. **R. V. Hogg and A. T. Craig** (1970). *Introduction to Mathematical Statistics.* Macmillan.

27. **R. Isaac** (1995). *The Pleasures of Probability.* Springer-Verlag.

28. **R. Jain** (1991). *The Art of Computer Systems Performance Analysis. Techniques for Experimental Design, Measurement, Simulation, and Modeling.* John Wiley & Sons.

29. **B. R. James** (1981). *Probabilidade: um curso em nível intermediário*. IMPA, CNPq.

30. **A. N. Kolmogorov** (1950). *Foundations of the Theory of Probability.* Chelsea Publishing Company.

31. **U. W. Kulisch and W. L. Miranker** (1981). *Computer Arithmetic in Theory and Practice.* New York, Academic Press.

32. **W. Kuo and M. J. Zuo** (2003). *Optimal Reliability Modeling Principles and Applications.* John Wiley & Sons.

33. **E. Lebensztayn e C. F. Coletti** (2008). *Notas de Aula.* Disponível em `http://www.ime.usp.br/~seccpg/mae`. Acessado em agosto de 2010.

34. **B. W. Lindgren** (1968). *Statistical Theory.* Macmillan.

35. **S. Lipschutz** (1993). *Probabilidade.* Coleção Schaum, 4a. Edição Revisada. Makrom Books.

36. **P. L. Meyer** (1983). *Probabilidade. Aplicações à Estatística.* Livros Técnicos e Científicos.

37. **D. C. Montgomery and G. C. Runger** (2003). *Estatística Aplicada e Probabilidade para Engenheiros.* Segunda Edição, Livros Técnicos e Científicos.

38. **R. E. Moore** (1966). *Interval Analysis.* Prentice Hall, Inc. Englewood Cliffs.

39. **R. E. Moore** (1979). *Methods and Applications of Interval Analysis.* SIAM, Philadelphia.

40. **F. Mosteller** (1965). *Fifty Challenging Problems in Probability.* Dover Publications.

41. **M. F. Neuts** (1995). *Algorithm Probability. A collection of problems.* Stochastic Modeling Series. ISBN 0 412 99791 X. Chapman & Hall.

42. **E. Parzen** (1960). *Modern Probability Theory and Its Applications.* John Wiley & Sons.

43. **E. Parzen** (1964). *Stochastic Processes.* Holden-Day. Second Edition.

44. **P. E. Pfeiffer** (1978). *Concepts of probability theory.* Second Revised Edition, Dover Publications, Inc.

45. **S. M. Ross** (1993). *Introduction to Probability Theory.* Academic Press.

46. **S. Russel and P. Norvig** (1995). *Artifitial Intelligence. A Modern Approach.* Prentice Hall.

47. **J. D. dos Santos e Z. C. da Silva** (2010). *Métodos Numéricos.* Editora Universitária/UFPE. 3a. Edição revisada.

48. **M. das Graças dos Santos** (2010). *Probabilidades Autovalidáveis para as Variáveis Aleatórias Exponencial, Normal e Uniforme.* Tese de Doutorado. Doutorado em Matemática Computacional, CCEN/CIn/UFPE.

49. **A. N. Shiryayev** (1984). *Probability.* Springer-Verlag, New York, Inc.

50. **F. Solomon** (1987). *Probability and Stochastic Processes.* Prentice-Hall, Inc.

51. **M. R. Spiegel** (1978). *Probabilidade e Estatística.* Coleção Schaum, McGraw.

52. **P. H. Sterbenz** (1974). *Floating-Point Computation.* Prentice-Hall, Englewood Cliffs, N. J.

53. **D. Stevenson**. (1981). *A Proposed Standard for Binary Floating-Point Arithmetic*. Computer, Vol. 14, No. 3, pp. 51-62.

54. **J. M. Stoyanov** (1997). *Counterexamples in Probability*. Second Edition. John Wiley & Sons.

55. **J. M. Stoyanov, I. Mirazchiiski, Z. Ignatov and M. Tamushev** (1989). *Exercise Manual in Probability Theory.* Kluwer Academic Publishers.

56. **A. A. Sveshnikov** (1978). *Problems in Probability Theory, Mathematical Statistics and Theory of Random Functions*. Dover.

57. **K. S. Trivedi** (1982). *Probability and Statistics with Reliability, Queueing and Computer Science Applications.* Prentice-Hall.

58. **H. C. Tuckwell** (1995). *Elementary Applications of Probability Theory. With an introdution to stochastic differential equations*. Chapman & Hall.

59. **J. S. Vandergraft** (1983). *Introduction to Numerical Computations*. Second Edition. Academic Press.

60. **N. A. Weiss and M. L. Yoseloff** (1975). *Matemática Finita*. Guanabara Dois.

61. **E. Wentzel and L. Ovcharov** (1986). *Applied Problems in Probability Theory.* Mir Publishers.

62. **J. H. Wilkinson** (1963). *Rounding Errors in Algebraic Processes.* Her Majesty's Stationery Office.

Apêndice A

Números de ponto-flutuante e Arredondamentos

Este apêndice refere-se à questão da representação dos números reais através de números de ponto-flutuante.

A.1 Números de ponto-flutuante

Todo número real x pode ser unicamente representado pela expansão b-ádica (Kullisch e Miranker 1979, 1981)

$$x = *d_n d_{n-1} \dots d_1 d_0 \cdot d_{-1} \dots = * \sum_{i=n}^{-\infty} d_i b^i \tag{1}$$

onde,

$$* \in \{+, -\},\ b \in \mathbb{N}, b > 1,$$
$$0 \leq d_i \leq b - 1,\ i = n(-1)(-\infty),$$
$$d_i \leq b - 2,\ \text{para uma quantidade infinita de índices } i.$$

Em (1) o ponto b-ádico pode ser mudado para qualquer posição desde que esta mudança seja convenientemente compensada. Se o ponto for mudado para a esquerda do primeiro dígito não nulo da expansão b-ádica do número x, através da multiplicação por uma correspondente potência de b, então tem-se o número x na forma *normalizada*. O número zero possui uma representação especial não normalizada.

A seguir uma explicação da necessidade da condição $d_i \leq b - 2$ para uma quantidade infinita de índices i.

$$0.3000 \dots = 3 \times 10^{-1} + 0 \times 10^{-2} + \dots = 3 \times 10^{-1} + \dots + 0 \times 10^{-n} + \dots$$

$$0.2999 \dots = 2 \times 10^{-1} + 9 \times 10^{-2} + \dots = 2 \times 10^{-1} + \dots + 9 \times 10^{-n} + \dots$$

Portanto as sequências

$$s_{1_n} = 3 \times 10^{-1} + \dots + 0 \times 10^{-n}$$

$$s_{2_n} = 2 \times 10^{-1} + \cdots + 9 \times 10^{-n}$$

são monótonas e limitadas, então convergem, e convergem para 0.3. Como pode ser visto em s_{2_n}, $d_i \not\leq b-2$ para uma quantidade infinita de índices i. Note que

$$s_{3_n} = 2 \times 10^{-1} + 8 \times 10^{-2} + \cdots + 8 \times 10^{-n}$$

converge para 0.29.

Seja R_b o conjunto de todos os x normalizados, mais o zero; em geral R_b não pode ser representado em computadores. Um subconjunto de R_b que é representável é definido a seguir.

Definição A.1.1: Um número real x é chamado um número de *ponto-flutuante normalizado* (Sterbenz 1974, Goldberg 1991, Forsythe 1977) ou um número de *ponto-flutuante* se

$$x = *0.d_1 d_2 \ldots d_l \times b^e \tag{2}$$

onde,

$$* \in \{+,-\},\ b \in \mathbb{N}, b > 1, \tag{3}$$

$$1 \leq d_1 \leq b-1, \tag{4}$$

$$0 \leq d_i \leq b-1,\ i = 2, \ldots, l, \tag{5}$$

$$e_{min} \leq e \leq e_{max}. \tag{6}$$

e, e_{min} e e_{max} são números inteiros, b é a *base* da representação, $*$ o sinal de x, $m = 0.d_1d_2\ldots d_l$ é a *mantissa* e e é o *expoente*. A mantissa tem comprimento l. Representa-se o número zero unicamente por

$$0 = 0.00\ldots0 \times b^{e_{min}}$$

$F(b, l, e_{min}, e_{max})$ é um *sistema de ponto-flutuante*, onde cada $x \in F$, $x \neq 0$, satisfaz, (2),(3),(4),(5),(6). Uma característica de F é que,

$$0 \in F,$$

$$1 \in F,$$

$$x \in F \Rightarrow -x \in F,\ \forall x \in F.$$

F é um conjunto finito. Tem exatamente $2(b-1)b^{l-1}(e_{max} - e_{min} + 1) + 1$ elementos, que só são igualmente espaçados entre sucessivas potências de b. Porque F é finito não há possibilidade de se representar o conjunto dos reais em detalhes. Como um modelo para $\mathbb{R}$, o conjunto F tem uma aritmética definida sobre ele. A questão é que esta aritmética não satisfaz propriedades básicas do ponto de vista da solução de problemas numéricos no conjunto dos reais. Supondo que x e y são números de ponto-flutuante, então nem sempre $x + y$ ou $x \cdot y$ estão em S. A manipulação algébrica de fórmulas é baseada em algumas leis fundamentais que são válidas nos reais, mas não em F; Goldberg (1991) e Sterbenz (1974) mostram que, considerando as operações aritméticas de adição e multiplicação em F, respectivamente, + e ·, são válidas (**V**) e não são válidas(**F**).

(**V**) $a + b = b + a,\ \forall a, b \in F,$
(**V**) $a \times b = b \times a,\ \forall a, b \in F,$
(**V**) $a + 0 = 0 + a = a,\ \forall a \in F,$
(**V**) $a \times 1 = 1 \times a = a,\ \forall a \in F,$
(**V**) $\forall a \in F,\ \exists(-a) \in F$ tal que $a + (-a) = (-a) + a = 0,$
(**F**) $(a + b) + c = a + (b + c),\ \forall a, b, c \in F,$
(**F**) $(a \times b) \times c = a \times (b \times c),\ \forall a, b, c \in F,$
(**F**) $a \times b = a \times c \Rightarrow b = c\ \forall a, b, c \in F,\ a \neq 0,$
(**F**) $a \times (b + c) = (a \times b) + (a \times c),\ \forall a, b, c \in F.$

Os primeiros computadores adotaram sistemas de representação de ponto-flutuante com diferentes bases, tamanho da palavra, número de dígitos significativos, etc. Este fato ocasionava sérios problemas tais como dificuldade de raciocinar em torno de provas relacionadas com algoritmos que geravam números de ponto-flutuante, porque os resultados aritméticos eram dependentes da máquina. Além disso, a ausência de um padrão que especificasse de modo detalhado as operações básicas e os formatos de dados acarretava a falta de portabilidade dos softwares. O esforço para produzir um padrão para ponto-flutuante originou o IEEE 754[1] (Cody 1988, Goldberg 1991, Stevenson 1991).

Exemplo A.1.2: Seja $F(b, l, e_{min}, e_{max})$.

(a) Quantos expoentes F tem, ou qual é a excursão do expoente?

$$e_{max} - e_{min} + 1$$

(b) Quantas são as mantissas positivas de F?

$$(b-1)b^{l-1}(e_{max} - e_{min} + 1)$$

(c) Qual o maior elemento de F, ou qual o máximo do conjunto F?

$$x_{max} = 0.(b-1)(b-1)\ldots(b-1) \times b^{e_{max}}$$

(d) Qual o menor elemento de F, ou qual o mínimo do conjunto F?

$$-x_{max}$$

(e) Qual o menor elemento positivo de F?

$$x_{min} = 0.(b-1)0\ldots0 \times b^{e_{min}}$$

(f) Qual o maior elemento negativo de F?

$$-x_{min}$$

[1]IEEE Standard for Binary Floating-Point Arithmetic

■

Este exemplo além de exibir elementos de F enfatiza a impossibilidade de F representar todos os números reais. Seja m $\mathcal{R} \subset \mathbb{R}$ o conjunto dos reais **representáveis** em qualquer computador onde os números reais sejam representados pelos números de ponto-flutuante. Portanto,

$$\mathcal{R} = [-x_{max}, -x_{min}] \cup \{0.0\ldots 0 \times b^{e_{min}}\} \cup [x_{min}, x_{max}],$$

onde

$$[-x_{max}, -x_{min}] = \{x \in \mathbb{R} : -x_{max} \leq x \leq -x_{min}\}$$

e

$$[x_{min}, x_{max}] = \{x \in \mathbb{R} : x_{min} \leq x \leq x_{max}\}.$$

Nem todos os elementos de $\mathcal{R}$ são representáveis em F e qualquer real fora de $\mathcal{R}$ é não representável em F. Os reais não representáveis estão nas regiões de *underflow*

$$(-x_{min}, 0.0\ldots 0 \times b^{e_{min}}) \cup (0.0\ldots 0 \times b^{e_{min}}, x_{min})$$

e *overflow*,

$$(-\infty, -x_{max}) \cup (x_{max}, +\infty)$$

Exemplo A.1.3: Seja $F(10, 3, -2, 2)$.

(a) Qual a representação do 0?

$$+0.000 \times 10^{-2}$$

(b)

$$x_{min} = 0.100 \times 10^{-2}$$

(c)

$$x_{max} = 0.999 \times 10^{2}$$

(d) Qual a representação do 1?

$$0.100 \times 10^{1}$$

(f) Quantos expoentes entre e inclusive os expoentes -2 e 1?

$$1 - (-2) + 1 = 4$$

(g) Quantas mantissas entre e inclusive os expoentes -2 e 1?

$$(2 - (-2)) \times 9 \times 10^{2} + 1 = 3601$$

Quantos elementos entre e inclusive os expoentes -2 e 1?

$$(2 - e_{min})(b - 1)b^{l-1} = 3600$$

■

A.2 Arredondamentos

Arredondar um número real x tem o significativo prático de transformar x em um número de F. Esta transformação é realizada através de uma função denominada de arredondamento.

Definição A.2.1: Seja $F = F(b, l, e_{min}, e_{max})$. Um arredondamento é uma função $\mathcal{O}$ definida por

$$\mathcal{O} : \mathcal{R} \to F$$

$$\mathcal{O}x = x_F$$

satisfazendo aos seguintes axiomas:

(i)

$$\mathcal{O}x = x, x \in F.$$

Isto é, o arredondamento não altera os elementos de F.

(ii)

$$x \leq y \Rightarrow \mathcal{O}x \leq \mathcal{O}y.$$

Portanto, a relação de ordem em $\mathcal{R}$ $(\leq)$ é preservada pela relação de ordem em F $(\leq)$.

(iii) Seja $x \in \mathcal{R}$, $x \notin F$ e $x_F = \mathcal{O}x$. Então, $\nexists\ y_F \in F$ entre x e x_F.

Este axioma afirma que x_F é o elemento de F mais próximo de $x \in \mathcal{R}$.

Os arredondamentos mais comuns são:

(i) $\bigtriangledown$, arredondamento para baixo ou em direção a $-\infty$,

$$\bigtriangledown : \mathcal{R} \to F$$

$$\forall x \in \mathcal{R}, \bigtriangledown x = \max\{y_F : y_F \leq x.\}$$

(ii) $\triangle$, arredondamento para cima ou em direção a $+\infty$,

$$\triangle : \mathcal{R} \to F$$

$$\forall x \in \mathcal{R}, \triangle x = \min\{y_F : y_F \geq x.\}$$

(iii) $\square$, arredondamento para o mais próximo ou arredondamento simétrico,

$$x \geq 0, \square x = \begin{cases} 0, x \in [0, b^{e_{min}-1}), \\ \bigtriangledown x, x \in [\bigtriangledown x, \frac{\bigtriangledown x + \triangle x}{2}), \\ \triangle x, x \in [\frac{\bigtriangledown x + \triangle x}{2}, \triangle x]. \end{cases}$$

$$x < 0, \square x = -\square(-x).$$

Convém enfatizar que $\bigtriangledown$, $\triangle$ e $\Box$ satisfazem aos axiomas da Definição A.2.1, e que $\bigtriangledown$ e $\triangle$ são exaustivamente usados na matemática intervalar e na aritmética de exatidão máxima (Moore (1966,1979), Kulisch e Miranker (1981)).

Santos e Silva (2010) provam que, fixado o expoente e, a distância entre dois números de ponto-flutuante consecutivos é b^{e-l+1}. Portanto, $\frac{1}{2}b^{e-l+1}$ é o erro de arredondamento cometido no arredondamento, especificamente o simétrico, é

$$\begin{aligned} \mid x - \Box x \mid &= 0, \ x = \Box x, \\ &\leq \frac{1}{2}b^{e-l+1}, \ x \neq \Box x. \end{aligned}$$

Os exemplos a seguir ilustram como realizar os arredondamentos $\bigtriangledown$, $\triangle$ e $\Box$.

Exemplo A.2.2: Seja $F(10, 4, -99, 99)$ e $x = 0.432987$. Portanto

$$\bigtriangledown x = 0.4329 \times 10^0, \ \triangle x = 0.4330 \times 10^0, \ \Box x = 0.4330 \times 10^0$$

■

Uma forma de computar $\Box x$ é encontrar os dois números de ponto-flutuante que contém x, x_{1_F} e x_{2_F} e atribuir a $\Box x$ o valor arredondado que apresentar a menor distância, em valor absoluto, a x.

Exemplo A.2.3: Seja $F(10, 4, -99, 99)$ e $x = 0.432987$. Portanto

$$x_{1_F} = 0.4329 \times 10^0 < 0.432987 < x_{2_F} = 0.4330 \times 10^0,$$

logo, $\mid 0.4329 - 0.432987 \mid = 0.000087$ e $\mid 0.4330 - 0.432987 \mid = 0.000013$, então $\Box x = 0.4330 \times 10^0$. ■

Exemplo A.2.4: Seja $F(10, 4, -99, 99)$ e $x = 0.34835$.

$$\bigtriangledown x = 0.3493, \ \triangle x = 0.3494, \ \Box x = 0.3494$$

■

Exemplo A.2.5: Este exemplo é só sobre o arredondamento simétrico. A primeira coluna do quadro abaixo contém elementos de $\mathcal{R}$, como, por exemplo, seriam digitados numa calculadora ou num computador digital. Na segunda coluna o elemento é posto no formato normalizado enfatizando quais dígitos vão ser descartados, na terceira coluna tem-se o valor em F e por último, como, usualmente resultados são exibidos.

$x \in \mathcal{R}$		$\Box x$	
232.489	$= 0.2324\underbrace{89}\times 10^3$	$:= 0.2325 \times 10^3$	$= 232.5$
16.17554	$= 0.1617\underbrace{554}\times 10^2$	$:= 0.1618 \times 10^2$	$= 16.18$
2.43456	$= 0.2434\underbrace{56}\times 10^1$	$:= 0.2435 \times 10^1$	$= 2.435$
2.43556	$= 0.2435\underbrace{56}\times 10^1$	$:= 0.2436 \times 10^1$	$= 2.436$
2.438556	$= 0.2438\underbrace{556}\times 10^1$	$:= 0.2439 \times 10^1$	$= 2.439$
0.34935	$= 0.3493\underbrace{5}\times 10^0$	$:= 0.3494 \times 10^0$	$= 0.3494$
0.00012467	$= 0.1246\underbrace{7}\times 10^{-3}$	$:= 0.1247 \times 10^{-3}$	$= 0.0001247$

■

Para concluir, enfatiza-se que a tarefa de representar e operar com os números reais em computadores é árdua. Inicialmente vem a escolha da representação (ponto-flutuante) em seguida como implementar na máquina (processador aritmético) a representação escolhida.

Índice