

Mônica C. Sandoval, Silvia L. P. Ferrari e Heleno Bolfarine

**Estatística para Estudantes de Matemática**  
**III Colóquio de Matemática da Região Sul**

**Florianópolis, SC**

**2014**



Mônica C. Sandoval, Silvia L. P. Ferrari e Heleno Bolfarine

**Estatística para Estudantes de Matemática**  
**III Colóquio de Matemática da Região Sul**

Minicurso apresentado no III Colóquio de Matemática da Região Sul, realizado na Universidade Federal de Santa Catarina, em maio de 2014.

Florianópolis, SC

2014



# Prefácio

Este texto foi preparado para um mini-curso de mesmo título no III Colóquio de Matemática da Região Sul. O mini-curso tem como objetivos apresentar conceitos básicos da inferência estatística e mostrar ao estudante de Matemática que a área de Estatística pode ser interessante para a continuação de seus estudos em nível de pós-graduação.

O texto é baseado no livro “Introdução à Inferência Estatística” dos dois primeiros autores. Selecionamos alguns tópicos capazes de ilustrar alguns conceitos centrais da teoria estatística. Incluímos também uma seção preliminar em que alguns conceitos de probabilidade são introduzidos para tornar o texto razoavelmente autocontido.

Agradecemos ao Comitê Científico e ao Comitê Local pela oportunidade de ministrar esse mini-curso e de divulgar a Estatística entre os estudantes de Matemática. Em particular, agradecemos ao Professor Artur Lopes pela sugestão de submeter este mini-curso à apreciação da organização do Colóquio.

São Paulo, fevereiro de 2014

Os autores

# Sumário

	<b>Introdução</b> . . . . .	<b>7</b>
<b>1</b>	<b>Conceitos Básicos</b> . . . . .	<b>9</b>
1.1	Preliminares . . . . .	9
1.2	Modelos estatísticos . . . . .	18
1.3	Problemas estatísticos . . . . .	21
1.4	Amostras, estatísticas e estimadores . . . . .	23
<b>2</b>	<b>Estimação Pontual</b> . . . . .	<b>27</b>
2.1	Estatísticas Suficientes . . . . .	27
2.2	Propriedades de estimadores . . . . .	33
2.2.1	Estimador não viciado e erro quadrático médio	34
2.2.2	Estimadores eficientes . . . . .	37
2.2.3	Estimadores Baseados em Estatísticas Suficientes	40
2.2.4	Estimadores consistentes . . . . .	42
2.3	Estimadores de máxima verossimilhança . . . . .	42
2.3.1	Propriedades dos Estimadores de Máxima Verossimilhança . . . . .	45
2.3.2	O Caso Multiparamétrico . . . . .	48
<b>3</b>	<b>Estimação Intervalar</b> . . . . .	<b>51</b>
3.1	Método da quantidade pivotal . . . . .	51
3.2	Intervalos para Populações Normais . . . . .	54
3.2.1	O caso de uma única amostra . . . . .	54
3.2.2	Duas amostras independentes . . . . .	55
3.3	Intervalos de Confiança Aproximados . . . . .	57

**Considerações Finais e Notas Bibliográficas . 59**

**Referências . . . . . 61**





# Introdução

Estatística é uma disciplina muito ampla com aplicações nas mais diversas áreas. Engloba o planejamento de experimentos, a obtenção, organização, resumo e visualização de dados, bem como a chamada inferência estatística.

Inferência estatística, que é o foco deste texto, compreende o uso de dados amostrais para se chegar a conclusões acerca de alguma característica da população, real ou hipotética, da qual os dados foram extraídos. Há diversos enfoques utilizados em inferência estatística, sendo os principais o enfoque frequentista, ou clássico, e o bayesiano.

Este texto apresenta alguns conceitos fundamentais da inferência estatística. Não procura ser abrangente pois foi escrito para servir de base a um mini-curso de quatro horas. Dada a limitação de tempo, foram escolhidos alguns tópicos e outros, embora relevantes, foram omitidos. Em particular, deve ser ressaltado que o texto trata apenas do enfoque frequentista.

O material apresentado a seguir está organizado da seguinte forma. No Capítulo 1, fazemos uma breve revisão de conceitos básicos de probabilidade que serão utilizados ao longo do texto. Os conceitos de modelos e problemas estatísticos, amostras, estatísticas e estimadores são também introduzidos. Os Capítulos 2 e 3 tratam de estimação pontual e intervalar respectivamente. Testes de hipóteses não são tratados neste livro. O livro se encerra com algumas considerações finais e notas bibliográficas.



# 1 Conceitos Básicos

## 1.1 Preliminares

Na prática, é comum estarmos diante de situações ou experimentos que envolvem incertezas. Experimentos que, ao serem repetidos nas mesmas condições, não produzem necessariamente o mesmo resultado são denominados experimentos aleatórios. O conjunto de todos os resultados possíveis de um experimento aleatório é chamado de espaço amostral ( $\Omega$ ), que pode ser finito ou infinito. O espaço amostral é infinito e enumerável se tiver um número infinito de elementos e puder ser colocado em correspondência biunívoca com o conjunto dos números naturais. É infinito não enumerável se não é nem finito e nem infinito enumerável.

**Exemplo 1.1.1.** Uma moeda é lançada duas vezes. Em cada lançamento pode ocorrer cara ( $c$ ) ou coroa ( $\bar{c}$ ). O espaço amostral é o conjunto  $\Omega = \{cc, c\bar{c}, \bar{c}c, \bar{c}\bar{c}\}$  e é finito.

**Exemplo 1.1.2.** Uma moeda é lançada sucessivamente até que apareça cara pela primeira vez. O espaço amostral é o conjunto  $\Omega = \{cc, \bar{c}c, \bar{c}\bar{c}c, \dots\}$  e é infinito enumerável.

**Exemplo 1.1.3.** Escolhemos uma lâmpada de um processo de fabricação e observamos o tempo até queimar. O espaço amostral é o conjunto dos números reais não negativos, ou seja,  $\Omega = \{x : x \in \mathbb{R}, x \geq 0\}$  e é infinito não enumerável.

Subconjuntos  $A, B, C, \dots$  do espaço amostral são cha-

mados de eventos e é de interesse calcular a “chance” ou probabilidade de ocorrência desses eventos. Uma função  $P$  que atribui valores numéricos aos eventos de um espaço amostral é chamada de probabilidade se satisfaz as seguintes condições:

- (a)  $0 \leq P[A] \leq 1$ , para todo  $A \subset \Omega$ ;
- (b)  $P[\Omega] = 1$ ;
- (c)  $P[\cup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i]$ , para todo  $A_i \cap A_{i'} = \emptyset$ ,  $i \neq i'$ .

Dados dois eventos  $A$  e  $B$  de um mesmo espaço amostral com  $P[B] > 0$ , a probabilidade condicional de  $A$  dado que ocorreu  $B$  é denotada por  $P[A|B]$  e definida por

$$P[A|B] = \frac{P[A \cap B]}{P[B]}.$$

Dois eventos são ditos independentes se a ocorrência de  $B$  não altera a probabilidade de ocorrência de  $A$ , isto é, se  $P[A|B] = P[A]$  ou, equivalentemente, se  $P[A \cap B] = P[A]P[B]$ .

Em alguns experimentos os resultados são numéricos, no entanto, em outros os elementos do espaço amostral não são expressos numericamente. Nesse caso, é conveniente associar valores numéricos a cada ponto de  $\Omega$ . Denomina-se variável aleatória qualquer função  $X$  do espaço amostral  $\Omega$  na reta real. No Exemplo 1.1.1, o interesse pode ser o número de caras que ocorre nos dois lançamentos da moeda. Então, define-se a função  $X$  que associa a cada ponto do espaço amostral o número de caras. Logo,  $X$  assume valores no conjunto  $\{0, 1, 2\}$ .

Qualquer variável aleatória  $X$  pode ser caracterizada por meio da função de distribuição (f.d.) definida por

$$F(x) = P[X \leq x], \quad x \in \mathfrak{R}.$$

Toda f.d. satisfaz as seguintes condições:

- (a)  $\lim_{x \rightarrow -\infty} F(x) = 0$  e  $\lim_{x \rightarrow \infty} F(x) = 1$ ;
- (b)  $F(x)$  é contínua à direita;
- (c)  $F(x)$  é não decrescente.

A partir da f.d. as variáveis aleatórias podem ser classificadas em discretas e contínuas. As variáveis aleatórias que assumem valores em um conjunto enumerável são denominadas discretas. Para essas variáveis a f.d. é descontínua e tem a forma de escada. Os pontos de descontinuidade  $x_i$ , para  $i = 1, 2, \dots$ , são os valores assumidos pela variável. A função de probabilidade (f.p.) de uma variável aleatória discreta é uma função,  $f$  digamos, que atribui probabilidade a cada um dos valores assumidos pela variável, ou seja,

$$f(x_i) = P[X = x_i], \quad i = 1, 2, 3, \dots$$

Toda f.p. deve satisfazer:

- (a)  $0 \leq f(x_i) \leq 1$ , para todo  $i$ ;
- (b)  $\sum_{i=1}^{\infty} f(x_i) = 1$ .

A partir da f.d. de uma variável aleatória discreta pode-se obter a f.p. como  $f(x_i) = P[X = x_i] = F(x_i) - F(x_i^-)$ , em que  $F(x_i^-)$  representa o limite de  $F(x)$  com  $x$  tendendo a  $x_i$  pela esquerda.

As variáveis que podem assumir todos os valores em um intervalo da reta (limitado ou não limitado) são denominadas contínuas. Uma variável será contínua se existir uma função  $f$ , denominada função densidade de probabilidade (f.d.p.), tal que

$$F(x) = \int_{-\infty}^x f(w)dw, \quad \text{para todo } x \in \mathfrak{R}.$$

As propriedades da f.d.p. são análogas às da f.p., ou seja,

- (a)  $f(x) \geq 0$ , para todo  $x$ ;

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

As variáveis aleatórias discretas e contínuas são caracterizadas pela função de probabilidade e pela função densidade de probabilidade, respectivamente. Quando for relevante explicitar o parâmetro  $\theta$  da distribuição, denotaremos a f.p. e a f.d.p por  $f(x|\theta)$ . Define-se como suporte da variável aleatória  $X$ , o conjunto dos números  $x$  tais que  $f(x|\theta) > 0$ .

Muitas vezes, quando um experimento aleatório é realizado, várias variáveis podem ser de interesse. Denomina-se vetor aleatório ou variável aleatória  $n$ -dimensional qualquer função  $\mathbf{X} = (X_1, \dots, X_n)$  de  $\Omega$  em  $\mathfrak{R}^n$ . A função de distribuição conjunta de  $\mathbf{X}$  é definida por

$$F(\mathbf{x}) = F(x_1, \dots, x_n) = P[X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n],$$

para todo  $\mathbf{x} = (x_1, \dots, x_n) \in \mathfrak{R}^n$ .

Denomina-se vetor aleatório discreto, o vetor aleatório  $\mathbf{X}$  cujos componentes  $X_1, \dots, X_n$  são variáveis aleatórias discretas. A função de probabilidade conjunta de  $\mathbf{X}$  é dada por

$$f(\mathbf{x}) = P[X_1 = x_1, X_2 = x_2, \dots, X_n = x_n]$$

e a função de probabilidade marginal de cada  $X_i$  é dada por

$$f_{X_i}(x_i) = P[X_i = x_i] = \sum_{x_j, j \neq i} P[X_1 = x_1, \dots, X_n = x_n].$$

Se os componentes são variáveis aleatórias contínuas, denomina-se o vetor  $\mathbf{X}$  de vetor aleatório contínuo. Um vetor aleatório é contínuo se existe uma função  $f : \mathfrak{R}^n \rightarrow \mathfrak{R}^+$  tal que

$$F(\mathbf{x}) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(\mathbf{y}) dy_1 dy_2 \dots dy_n.$$

A função  $f$  é denominada função densidade de probabilidade conjunta. A função densidade de probabilidade marginal de cada  $X_i$  é dada por

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(\mathbf{y}) dy_1 \dots dy_{i-1} dy_{i+1} \dots dy_n.$$

As propriedades do vetor aleatório discreto ou contínuo são análogas às dos respectivos casos unidimensionais.

Sejam  $X_1$  e  $X_2$  duas variáveis aleatórias discretas (contínuas) com f.p. (f.d.p.) conjunta  $f_{X_1, X_2}(x_1, x_2)$ . Seja  $x_1$  um ponto no suporte de  $X_1$ , ou seja,  $x_1$  é tal que  $f_{X_1}(x_1) > 0$ . A função de probabilidade condicional de  $X_2$  dado  $X_1 = x_1$  é definida por

$$f_{X_2|X_1=x_1}(x_2) = \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_1}(x_1)}.$$

Um conceito importante é o de independência entre variáveis aleatórias. As variáveis aleatórias  $X_1, X_2, \dots, X_n$  são independentes se e somente se para todo  $(x_1, x_2, \dots, x_n)$ ,

$$f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2) \dots f(x_n).$$

As principais medidas utilizadas para resumir o comportamento de uma variável aleatória são a média e a variância. Para variáveis aleatórias discretas, o valor esperado, esperança matemática ou média de  $X$  é dada por

$$E[X] = \mu = \sum_{i=1}^{\infty} x_i P[X = x_i].$$

Para variáveis aleatória contínuas  $X$  com f.d.p.  $f(x)$  o valor esperado é dado por

$$E[X] = \mu = \int_{-\infty}^{\infty} x f(x) dx.$$

A média de uma função  $g(X)$  é dada por

$$E[g(X)] = \sum_{i=1}^{\infty} g(x_i)P[X = x_i]$$

no caso discreto e

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

no caso contínuo. A variância de uma variável aleatória  $X$  é definida por  $\text{Var}[X] = \sigma^2 = E[(X - E(X))^2]$  e a raiz quadrada da variância é denominada desvio padrão.

A esperança de uma variável aleatória dá idéia da posição da distribuição de probabilidade dessa variável e a variância fornece uma característica importante de uma variável aleatória que é a sua variabilidade, avaliada pela dispersão de seus valores em relação à sua média. Uma expressão alternativa para obtenção da variância é  $\text{Var}[X] = E[X^2] - E[X]^2$ .

A seguir apresentamos algumas propriedades da média e da variância. Sendo  $a$  e  $b$  constantes quaisquer, temos

$$E[aX + b] = aE[X] + b \quad \text{e} \quad \text{Var}[aX + b] = a^2\text{Var}[X].$$

Podemos observar que as alterações feitas nos valores da variável se refletem na sua média. Já, em relação à variância, vemos que apenas a multiplicação por constante interfere na dispersão da variável. A variância não se altera com o acréscimo de constantes.

Para um vetor aleatório  $\mathbf{X} = (X_1, \dots, X_n)$  temos que

$$E[X_1 + \dots + X_n] = E[X_1] + \dots + E[X_n]$$

e, se as variáveis aleatórias forem independentes,

$$\text{Var}[X_1 + \dots + X_n] = \text{Var}[X_1] + \dots + \text{Var}[X_n].$$



Algumas distribuições de probabilidade importantes são apresentadas a seguir.

**Distribuição normal.** Dizemos que  $X$  tem distribuição normal com parâmetros  $\mu$  e  $\sigma^2$ , que denotamos por  $X \sim N(\mu, \sigma^2)$ , se a f.d.p. de  $X$  é dada por

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty,$$

em que  $-\infty < \mu < \infty$  e  $\sigma^2 > 0$ . Neste caso, o suporte de  $X$  é a reta real. A média e a variância de  $X$  são dadas por

$$E[X] = \mu \quad \text{e} \quad \text{Var}[X] = \sigma^2.$$

O cálculo de probabilidades com a f.d.p. normal não pode ser feito pela integral, mas podem ser obtidos numericamente. Probabilidades envolvendo a distribuição  $N(0, 1)$ , também chamada de distribuição normal padrão, são tabeladas. Uma vez que se  $X \sim N(\mu, \sigma^2)$  então  $Z = (X - \mu)/\sigma \sim N(0, 1)$ , temos

$$P[a \leq X \leq b] = P\left[\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right],$$

que podem ser calculadas usando uma tabela da distribuição normal padrão. A distribuição normal é comumente utilizada para descrever variáveis como peso, altura, pressão arterial, quociente de inteligência, etc.

**Distribuição exponencial.** Dizemos que  $X$  tem distribuição exponencial com parâmetro  $\theta$ , que denotamos por  $X \sim \text{Exp}(\theta)$ , se a f.d.p. de  $X$  é dada por

$$f(x|\theta) = \theta e^{-\theta x}, \quad x > 0,$$

em que  $\theta > 0$ . Neste caso, o suporte de  $X$  é  $A = \{x, x > 0\}$ . A média e a variância de  $X$  são dadas por

$$E[X] = \frac{1}{\theta} \quad \text{e} \quad \text{Var}[X] = \frac{1}{\theta^2}.$$

A distribuição exponencial é comumente empregada para descrever tempo de vida de equipamentos. A distribuição exponencial tem a bem conhecida propriedade da falta de memória, ou seja,  $P[X > s + t | X > s] = P[X > t]$ , para todo  $s, t \geq 0$ .

**Distribuição de Bernoulli.** Dizemos que a variável aleatória  $X$  tem distribuição de Bernoulli com parâmetro  $\theta$ , que denotamos por  $X \sim \text{Bernoulli}(\theta)$ , se sua f.p. é dada por

$$f(x|\theta) = \begin{cases} \theta, & \text{se } x = 1, \\ 1 - \theta, & \text{se } x = 0. \end{cases}$$

A média e a variância de  $X$  são dadas por

$$E[X] = \theta \quad \text{e} \quad \text{Var}[X] = \theta(1 - \theta).$$

Neste caso, o suporte de  $X$  é  $A = \{0, 1\}$ .

**Distribuição binomial.** Dizemos que a variável aleatória  $X$  tem distribuição binomial com parâmetros  $n$  e  $\theta$ , que denotamos por  $X \sim \text{Binomial}(n, \theta)$ , se sua f.p. é dada por

$$f(x|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, \dots, n,$$

em que  $0 < \theta < 1$ . Nesse caso, o suporte de  $X$  é  $A = \{0, 1, \dots, n\}$ . A média e a variância de  $X$  são dadas por

$$E[X] = n\theta \quad \text{e} \quad \text{Var}[X] = n\theta(1 - \theta).$$

Se  $X$  tem distribuição Binomial( $n, \theta$ ), podemos escrever  $X = X_1 + \dots + X_n$ , sendo  $X_1, \dots, X_n$  variáveis aleatórias independentes com distribuição Bernoulli( $\theta$ ). A distribuição binomial é comumente empregada em situações em que cada observação da amostra admite apenas dois resultados, sucesso ou fracasso, como, por exemplo, em pesquisas eleitorais, em que cada indivíduo é ou não favorável a determinado partido ou candidato, ou

ainda em controle de qualidade, em que cada peça produzida em uma linha de produção é ou não defeituosa.

**Distribuição de Poisson.** Dizemos que a variável aleatória  $X$  tem distribuição de Poisson com parâmetro  $\theta$ , que denotamos por  $X \sim \text{Poisson}(\theta)$ , se sua f.p. é dada por

$$f(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}, \quad x = 0, 1, \dots,$$

em que  $\theta > 0$ . Neste caso, o suporte de  $X$  é o conjunto dos números naturais. A média e a variância de  $X$  são

$$E[X] = \text{Var}[X] = \theta.$$

A distribuição de Poisson é utilizada para descrever variáveis que representam uma contagem como, por exemplo, o número de chamadas que chegam a uma central telefônica, o número de partículas  $\alpha$  emitidas por uma fonte radioativa ou o número de pessoas que chegam a determinada fila, sempre em um intervalo de tempo fixado.

**Distribuição uniforme.** Dizemos que  $X$  tem distribuição uniforme no intervalo  $[0, \theta]$ , que denotamos por  $X \sim U(0, \theta)$ , se a f.d.p. de  $X$  é dada por

$$f(x|\theta) = \frac{1}{\theta} I_{[0, \theta]}(x),$$

em que  $\theta > 0$ , e  $I_{[0, \theta]}(x)$  é a função indicadora do intervalo  $[0, \theta]$ , isto é,

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

Neste caso, o suporte de  $X$  é  $A = [0, \theta]$  e depende do parâmetro  $\theta$ . A média e a variância de  $X$  são

$$E[X] = \frac{\theta}{2} \quad \text{e} \quad \text{Var}[X] = \frac{\theta^2}{12}.$$

**Distribuição qui-quadrado.** Dizemos que  $X$  tem distribuição qui-quadrado com  $\nu$  graus de liberdade, que denotamos por  $X \sim \chi_\nu^2$ , se a f.d.p. de  $X$  é dada por

$$f(y|\nu) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} y^{\nu/2-1} e^{-y/2}, \quad y > 0,$$

em que  $\nu > 0$  e  $\Gamma(\cdot)$  representa a função gama. Neste caso, o suporte de  $X$  é  $A = \{x, x > 0\}$ . A média e a variância de  $X$  são dadas por

$$E[X] = \nu \quad \text{e} \quad \text{Var}[X] = 2\nu.$$

A distribuição  $\chi_\nu^2$  é tabelada para diferentes valores de  $\nu$ .

**Distribuição t de Student.** Dizemos que  $X$  tem distribuição t de Student com  $\nu$  graus de liberdade, que denotamos por  $X \sim t_\nu$ , se a f.d.p. de  $X$  é dada por

$$f(y|\nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \left(1 + \frac{y^2}{\nu}\right)^{-(\nu+1)/2}, \quad -\infty < y < \infty,$$

em que  $\nu > 0$ . Neste caso, o suporte de  $X$  é a reta real. A média e a variância de  $X$  são dadas por

$$E[X] = 0 \quad \text{e} \quad \text{Var}[X] = \frac{\nu}{\nu-2}.$$

A distribuição  $t_\nu$  é tabelada para diferentes valores de  $\nu$ .

## 1.2 Modelos estatísticos

Considere uma f.d.p. ou f.p.  $f(x|\theta)$  em que o parâmetro  $\theta \in \Theta \subset \mathfrak{R}^k$ . Cada valor fixado de  $\theta \in \Theta$  corresponde a uma distribuição de probabilidades e o conjunto  $\mathcal{F} = \{f(x|\theta), \theta \in \Theta\}$  define uma família de distribuições de probabilidade. Um modelo estatístico (paramétrico) para uma variável aleatória (ou

um vetor aleatório) observável  $X$  refere-se à suposição de que  $X$  tem distribuição na família  $\mathcal{F}$ .

**Definição 1.2.1.** *O conjunto  $\Theta$  em que  $\theta$  toma valores é denominado espaço paramétrico.*

A seleção do modelo hipotético a ser utilizado numa análise estatística não é uma tarefa trivial. É fundamental que o modelo represente, na medida do possível, a complexidade que envolve o mundo real da variável em estudo. Neste texto focaremos em alguns modelos simples, derivados de distribuições de probabilidade introduzidas na Seção 1.1, e que são comumente utilizados em análise de dados.

Os modelos normal, exponencial, Bernoulli, binomial e Poisson são membros de uma família de modelos chamada de família exponencial.

**Definição 1.2.2.** *Dizemos que a distribuição da variável aleatória (ou do vetor aleatório)  $X$  pertence à família exponencial unidimensional de distribuições se sua f.p. ou f.d.p. é dada por*

$$f(x|\theta) = \exp\{c(\theta)T(x) + d(\theta) + S(x)\}, \quad x \in A, \quad (1.1)$$

em que  $c(\cdot)$  e  $d(\cdot)$  são funções reais de  $\theta$ ,  $T(\cdot)$  e  $S(\cdot)$  são funções reais de  $x$  e o conjunto  $A$  não depende de  $\theta$ .

No caso em que  $X$  é contínua, para que  $f(x|\theta)$  em (1.1) seja uma f.d.p. é necessário que

$$\int_A \exp\{c(\theta)T(x) + d(\theta) + S(x)\}dx = 1,$$

ou seja,

$$\int_A \exp\{c(\theta)T(x) + S(x)\}dx = \exp\{-d(\theta)\},$$

de modo que  $d(\theta)$  está associado à constante de normalização da densidade. Resultado similar vale para o caso em que  $X$  é uma variável aleatória discreta.

**Exemplo 1.2.1.** Seja  $X \sim \text{Bernoulli}(\theta)$ . Podemos escrever<sup>1</sup>

$$f(x|\theta) = \theta^x (1 - \theta)^{1-x} = \exp \left\{ x \log \left( \frac{\theta}{1 - \theta} \right) + \log(1 - \theta) \right\},$$

$x \in \{0, 1\}$ . Portanto, a distribuição de  $X$  pertence à família exponencial unidimensional com

$$c(\theta) = \log \left( \frac{\theta}{1 - \theta} \right), \quad d(\theta) = \log(1 - \theta),$$

$$T(x) = x, \quad S(x) = 0, \quad A = \{0, 1\}.$$

**Exemplo 1.2.2.** Seja  $X \sim N(\mu, 1)$ . Temos que

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu)^2}{2} \right\} = \exp \left\{ \mu x - \frac{\mu^2}{2} - \frac{x^2}{2} - \log \sqrt{2\pi} \right\}.$$

Portanto, a distribuição da variável aleatória  $X$  pertence à família exponencial unidimensional com

$$c(\mu) = \mu, \quad d(\mu) = -\frac{\mu^2}{2},$$

$$T(x) = x, \quad S(x) = -\frac{x^2}{2} - \log \sqrt{2\pi}, \quad A = \mathfrak{R}.$$

**Definição 1.2.3.** Dizemos que a distribuição da variável aleatória (ou do vetor aleatório)  $X$  pertence à família exponencial de dimensão  $k$  se a f.d.p. ou a f.p. de  $X$  é dada por

$$f(x|\theta) = \exp \left\{ \sum_{j=1}^k c_j(\theta) T_j(x) + d(\theta) + S(x) \right\}, \quad x \in A, \quad (1.2)$$

<sup>1</sup> Neste texto,  $\log$  denota logaritmo na base  $e$ .

em que  $c_j(\cdot)$  e  $d(\cdot)$  são funções reais de  $\theta$  e  $T_j(\cdot)$  e  $S(\cdot)$  são funções reais de  $x$ , para  $j = 1, \dots, k$  e  $A$  não depende de  $\theta$ . Como no caso unidimensional,  $d(\theta)$  está associado à constante de normalização de (1.2).

**Exemplo 1.2.3.** Seja  $X \sim N(\mu, \sigma^2)$ . Temos que

$$\begin{aligned} f(x|\theta) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ &= \exp \left\{ -\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2} - \frac{1}{2} \log \sigma^2 - \log \sqrt{2\pi} \right\}, \end{aligned}$$

Portanto, a distribuição da variável aleatória  $X$  pertence à família exponencial bidimensional com

$$\begin{aligned} T_1(x) &= x, & T_2(x) &= x^2, & c_1(\theta) &= \frac{\mu}{\sigma^2}, & c_2(\theta) &= -\frac{1}{2\sigma^2}, \\ d(\theta) &= -\frac{\mu^2}{2\sigma^2} - \frac{1}{2} \log \sigma^2, & S(x) &= -\log \sqrt{2\pi}, & A &= \mathfrak{R}. \end{aligned}$$

### 1.3 Problemas estatísticos

Inferência estatística compreende o uso de dados amostrais, digamos  $\mathbf{x} = (x_1, \dots, x_n)$ , para se chegar a conclusões acerca de algum aspecto da população (real ou hipotética) da qual os dados foram extraídos. Em inferência estatística paramétrica, de que este livro trata, os dados são modelados como valores observados de variáveis aleatórias, colecionadas no vetor aleatório  $\mathbf{X} = (X_1, \dots, X_n)$ , que seguem algum modelo estatístico pré-estabelecido que depende de um parâmetro  $\theta \in \Theta$  ( $\theta$  pode ser escalar ou vetorial). Desta forma, a distribuição de  $X$  é parcialmente conhecida, já que a forma de sua f.d.p. ou f.p.,  $f(\mathbf{x}|\theta)$  digamos, é conhecida, mas seu parâmetro é desconhecido.

**Definição 1.3.1.** *Seja  $X$  uma variável aleatória (ou um vetor aleatório) com f.d.p. ou f.p.  $f(x|\theta)$  pertencente à família*

$\mathcal{F} = \{f(x|\theta), \theta \in \Theta\}$ , em que  $\theta$  é um parâmetro desconhecido e  $\Theta \subset \mathbb{R}^k$ . Chamamos de *inferência estatística* o problema que consiste em especificar um ou mais valores para  $\theta$ , baseado em um conjunto de valores observados de  $X$ .

Vamos assumir que a família  $\mathcal{F}$  é *identificável*, ou seja, cada elemento de  $\mathcal{F}$  é unicamente identificado pelo valor de  $\theta$ .

Três problemas fundamentais em inferência estatística são: estimação pontual, estimação intervalar e testes de hipóteses. Num problema de **estimação pontual** o objetivo é procurar, segundo algum critério especificado, um valor que represente adequadamente o parâmetro desconhecido  $\theta$ , ou um ou mais de seus componentes. **Estimação intervalar**, por outro lado, busca estimar os componentes de  $\theta$  através de intervalos de valores plausíveis. Finalmente, em problemas de **testes de hipóteses**, o objetivo é verificar a veracidade de afirmações sobre os parâmetros desconhecidos.

Por exemplo, uma máquina de empacotamento automático de açúcar é regulada para produzir pacotes de 1 kg, em média, com desvio padrão de 0,02 kg. Suponhamos que, por experiência passada, seja razoável supor que o peso dos pacotes produzidos pela máquina seguem uma distribuição normal. Como é possível que a máquina se desregule, é de interesse inferir sobre o peso médio dos pacotes ( $\mu$  digamos). Se uma amostra de pacotes é aleatoriamente selecionada da produção, o peso médio pode ser estimado através da média amostral dos pesos dos pacotes selecionados. Ou ainda, é possível construir um intervalo de valores plausíveis para o peso médio. Em geral, o interesse é saber se este é, de fato, igual a 1 kg. Neste caso, um teste da hipótese  $H_0 : \mu = 1$  contra a hipótese  $H_1 : \mu \neq 1$  pode ser



construído.

No presente texto, abordaremos exclusivamente problemas de estimação, tanto pontual quanto intervalar.

## 1.4 Amostras, estatísticas e estimadores

Nesta seção os conceitos de amostra aleatória, estatística e estimador são formalizados.

**Definição 1.4.1.** *O conjunto de valores de uma característica (observável) associada a uma coleção de indivíduos ou objetos de interesse é dito ser uma população.*

Qualquer parte (ou subconjunto) de uma população é denominada uma amostra. Formalmente, temos a seguinte definição.

**Definição 1.4.2.** *Seja  $X$  uma variável aleatória com f.d.p. ou f.p.  $f(x|\theta)$ . Um vetor  $(X_1, \dots, X_n)$  de  $n$  variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com com f.d.p. ou f.p.  $f(x|\theta)$  é dito ser uma amostra aleatória (ordenada) de tamanho  $n$  da distribuição de  $X$ . Nesse caso, a f.d.p. ou f.p. conjunta de  $X_1, \dots, X_n$  é dada por*

$$f(x_1, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta) = f(x_1|\theta) \dots f(x_n|\theta).$$

**Definição 1.4.3.** *Qualquer função da amostra que não dependa de parâmetros desconhecidos é denominada uma estatística.*

**Exemplo 1.4.1.** Seja  $(X_1, \dots, X_n)$  uma amostra aleatória da variável aleatória  $X$ . Exemplos de estatísticas são

$$X_{(1)} = \min(X_1, \dots, X_n), \quad X_{(n)} = \max(X_1, \dots, X_n),$$

$$\tilde{X} = \text{med}(X_1, \dots, X_n), \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2,$$

em que  $\min(\cdot)$ ,  $\max(\cdot)$  e  $\text{med}(\cdot)$  denotam, respectivamente, o mínimo, o máximo e a mediana amostrais. Note que  $\bar{X}$  e  $\hat{\sigma}^2$  denotam, respectivamente, a média e a variância amostrais.

**Definição 1.4.5.** *Qualquer estatística que assume valores somente no espaço paramétrico  $\Theta$  é um estimador de  $\theta$ .*

Em outras palavras, um estimador de  $\theta$  é qualquer função da amostra,  $\hat{\theta} = \hat{\theta}(\mathbf{X})$  digamos, que assume valores apenas em  $\Theta$ , e é usada para estimar o parâmetro desconhecido. O valor do estimador calculado na amostra observada  $\mathbf{x}$ , ou seja  $\hat{\theta}(\mathbf{x})$ , é chamado de estimativa de  $\theta$ .

Em muitas situações, o interesse é estimar uma função  $g(\theta)$ . Por exemplo, em um modelo normal, que é especificado por dois parâmetros, média ( $\mu$ ) e variância ( $\sigma^2$ ), o interesse do pesquisador pode estar focado na média e, neste caso,  $\sigma^2$  é visto como um parâmetro de perturbação, ou seja, necessário à especificação do modelo mas não de interesse. Aqui,  $\theta = (\mu, \sigma^2)$  e  $g(\theta) = \mu$ .

**Definição 1.4.6.** *Qualquer estatística que assume valores somente no conjunto dos possíveis valores de  $g(\theta)$  é um estimador de  $g(\theta)$ .*

Finalizamos esta seção apresentando um teorema que fornece resultados utilizados em inferência para populações normais.

**Teorema 1.5.1.** *Seja  $(X_1, \dots, X_n)$  uma amostra aleatória de tamanho  $n$  da distribuição  $N(\mu, \sigma^2)$ . Então*

(i)  $\bar{X}$  e  $S^2$  são independentes,

$$(ii) \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2,$$

$$(iii) \frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1},$$

em que  $\bar{X} = \sum_{i=1}^n X_i/n$  e  $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ .



## 2 Estimação Pontual

### 2.1 Estatísticas Suficientes

Seja  $\mathbf{X} = (X_1, \dots, X_n)$  uma amostra aleatória da variável aleatória  $X$  com f.d.p ou f.p.  $f(x|\theta)$ . Inferências sobre o parâmetro  $\theta$  são feitas com base em alguma estatística, ou seja, uma função dos dados. Certas estatísticas guardam toda a informação que a amostra contém sobre o parâmetro. Em outras palavras, possibilitam o resumo dos dados sem perda de informação sobre  $\theta$ . Nesse caso, o conhecimento apenas da estatística (e não necessariamente da amostra completa) é suficiente para que sejam feitas inferências sobre  $\theta$ . A seguir apresentamos a definição formal de estatística suficiente.

**Definição 2.1.1.** *Dizemos que a estatística  $T = T(\mathbf{X})$  é suficiente para  $\theta$  se a distribuição condicional de  $\mathbf{X}$  dado  $T$  for independente de  $\theta$ .*

**Exemplo 2.1.1. Amostra aleatória de Bernoulli( $\theta$ ).** Seja  $(X_1, \dots, X_n)$  uma amostra aleatória da distribuição de Bernoulli( $\theta$ ). Verifiquemos se a estatística  $T = \sum_{i=1}^n X_i$  é suficiente para  $\theta$ . De acordo com a Definição 2.1.1,  $T$  é suficiente para  $\theta$  se a probabilidade condicional  $P[X_1 = x_1, \dots, X_n = x_n | T = t]$  for independente de  $\theta$ , para todo  $t = 0, 1, \dots, n$ . Temos, para  $x_i \in \{0, 1\}$ ,  $i = 1, \dots, n$ , e  $t = 0, \dots, n$ ,

$$P[X_1 = x_1, \dots, X_n = x_n | T = t] = \begin{cases} 0, & \text{se } \sum_{i=1}^n x_i \neq t, \\ \frac{P[X_1 = x_1, \dots, X_n = x_n]}{P[T=t]}, & \text{se } \sum_{i=1}^n x_i = t. \end{cases} \quad (2.1)$$

Se  $\sum_{i=1}^n x_i = t$ , temos

$$\begin{aligned} P[X_1 = x_1, \dots, X_n = x_n | T = t] &= \frac{P[X_1 = x_1] \dots P[X_n = x_n]}{\binom{n}{t} \theta^t (1 - \theta)^{n-t}} \\ &= \frac{\theta^{x_1} (1 - \theta)^{1-x_1} \dots \theta^{x_n} (1 - \theta)^{1-x_n}}{\binom{n}{t} \theta^t (1 - \theta)^{n-t}} = \frac{\theta^t (1 - \theta)^{n-t}}{\binom{n}{t} \theta^t (1 - \theta)^{n-t}} = \frac{1}{\binom{n}{t}}, \end{aligned}$$

pois  $X_1, \dots, X_n$  so independentes e  $T \sim \text{Binomial}(n, \theta)$ . Assim,

$$P[X_1 = x_1, \dots, X_n = x_n | T = t] = \begin{cases} 0, & \text{se } \sum_{i=1}^n x_i \neq t, \\ \frac{1}{\binom{n}{t}}, & \text{se } \sum_{i=1}^n x_i = t, \end{cases}$$

que no depende de  $\theta$ . Portanto,  $T = \sum_{i=1}^n X_i$   uma estatística suficiente para  $\theta$ .

**Exemplo 2.1.2. Amostra aleatria de Poisson( $\theta$ ).** Seja  $(X_1, \dots, X_n)$  uma amostra aleatria da distribuio de Poisson com parmetro  $\theta$ . Verifiquemos se  $T = \sum_{i=1}^n X_i$   suficiente para  $\theta$ .  possvel mostrar que  $T$  tem distribuio de Poisson com parmetro  $n\theta$ . Assim, para  $x_i, t = 0, 1, 2, \dots, i = 1, \dots, n$ , temos que  $P[X_1 = x_1, \dots, X_n = x_n | T = t]$   dada por (2.1) e, ento, se  $\sum_{i=1}^n x_i = t$ , temos

$$\begin{aligned} P[X_1 = x_1, \dots, X_n = x_n | T = t] &= \frac{P[X_1 = x_1] \dots P[X_n = x_n]}{P[T = t]} \\ &= \frac{t!}{x_1! \dots x_n! n^t}, \end{aligned}$$

que  independente de  $\theta$ . Portanto,  $\sum_{i=1}^n X_i$   uma estatística suficiente para  $\theta$ .

A Definio 2.1.1 apenas permite verificar se determinada estatística  ou no suficiente. Contudo, no pode ser utilizada como um mtodo para obteno de estatísticas suficientes. Um procedimento para a obteno de estatísticas suficientes  o

critério da fatoração. Antes de introduzi-lo é conveniente abordar o conceito de função de verossimilhança.

**Definição 2.1.2.** *Seja  $\mathbf{X} = (X_1, \dots, X_n)$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$  com f.d.p. ou f.p.  $f(x|\theta)$ ,  $\theta \in \Theta$ , em que  $\Theta$  é o espaço paramétrico. A função de verossimilhança de  $\theta$  correspondente à amostra aleatória observada  $\mathbf{x} = (x_1, \dots, x_n)$  é dada por*

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i|\theta), \quad \theta \in \Theta. \quad (2.2)$$

Note que a função de verossimilhança é a f.d.p. ou f.p. conjunta de  $\mathbf{X}$  avaliada na amostra observada  $\mathbf{x}$ , vista como função do parâmetro  $\theta$ .

**Teorema 2.1.1.** *(Critério da Fatoração de Neyman) Seja  $\mathbf{X} = (X_1, \dots, X_n)$  uma amostra aleatória da distribuição da variável aleatória  $X$  com f.d.p. ou f.p.  $f(x|\theta)$  e seja  $L(\theta; \mathbf{x})$  a função de verossimilhança. A estatística  $T = T(\mathbf{X})$  é suficiente para  $\theta$  se e somente se*

$$L(\theta; \mathbf{x}) = h(\mathbf{x})g_\theta(T(\mathbf{X})), \quad \theta \in \Theta, \quad (2.3)$$

em que  $h(\mathbf{x})$  é uma função que depende apenas de  $\mathbf{x}$  (não depende de  $\theta$ ) e  $g_\theta(T(\mathbf{x}))$  depende de  $\theta$  e depende de  $\mathbf{x}$  somente através de  $T$ .

**Prova.** Provemos o teorema apenas para o caso discreto, em que  $L(\theta; \mathbf{x}) = P_\theta[\mathbf{X} = \mathbf{x}]$ . Suponhamos que (2.3) esteja verificada. Então,

$$P_\theta[\mathbf{X} = \mathbf{x}] = h(\mathbf{x})g_\theta(T(\mathbf{x})).$$

Como

$$P[\mathbf{X} = \mathbf{x}|T(\mathbf{X}) = t] = \begin{cases} 0, & \text{se } T(\mathbf{x}) \neq t \\ \frac{P_\theta[\mathbf{X}=\mathbf{x}, T(\mathbf{X})=t]}{P_\theta[T(\mathbf{X})=t]}, & \text{se } T(\mathbf{x}) = t, \end{cases}$$

temos que, quando  $T(\mathbf{x}) \neq t$ ,  $P_\theta[\mathbf{X} = \mathbf{x}|T(\mathbf{x}) = t] = 0$ , que é independente de  $\theta$  e, portanto, (2.3) est verificada. Quando  $T(\mathbf{x}) = t$ ,

$$\begin{aligned} P_\theta[\mathbf{X} = \mathbf{x}|T(\mathbf{X}) = t] &= \frac{P_\theta[\mathbf{X} = \mathbf{x}, T(\mathbf{X}) = t]}{P_\theta[T = t]} = \frac{P_\theta[\mathbf{X} = \mathbf{x}]}{P_\theta[T = t]} \\ &= \frac{h(\mathbf{x})g_\theta(t)}{\sum_{\{\mathbf{x}; T(\mathbf{x})=t\}} h(\mathbf{x})g_\theta(t)} = \frac{h(\mathbf{x})}{\sum_{\{\mathbf{x}; T(\mathbf{x})=t\}} h(\mathbf{x})}, \end{aligned}$$

que é independente de  $\theta$  e, portanto,  $T = T(\mathbf{X})$  é uma estatística suficiente para  $\theta$ .

Suponhamos agora que  $T = T(\mathbf{X})$  seja uma estatística suficiente. Sendo  $T(\mathbf{x}) = t$ , temos que

$$\begin{aligned} P_\theta[\mathbf{X} = \mathbf{x}] &= P_\theta[\mathbf{X} = \mathbf{x}, T(\mathbf{x}) = t] \\ &= P_\theta[\mathbf{X} = \mathbf{x}|T(\mathbf{x}) = t]P_\theta[T(\mathbf{X}) = t] = h(\mathbf{x})g_\theta(t), \end{aligned}$$

pois a distribuio condicional de  $\mathbf{X}$  dado  $T$  independente de  $\theta$ .

**Exemplo 2.1.3. Amostra aleatria de Poisson( $\theta$ ).** Para  $\mathbf{x} =$

$(x_1, \dots, x_n)$  com  $x_i = 1, 2, \dots, i = 1, \dots, n$ , temos

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x_i!} = \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}, \quad \theta > 0.$$

Portanto, tomando  $T(\mathbf{x}) = \sum_{i=1}^n x_i$ ,

$$h(\mathbf{x}) = \frac{1}{\prod_{i=1}^n x_i!} \quad \text{e} \quad g_\theta(T(\mathbf{x})) = e^{-n\theta} \theta^{\sum_{i=1}^n x_i},$$

temos, pelo critrio da fatoraco, que  $T(\mathbf{X}) = \sum_{i=1}^n X_i$  é uma estatística suficiente para  $\theta$ .



**Exemplo 2.1.4. Amostra aleatória de  $U(0, \theta)$ .** Seja  $\mathbf{X} = (X_1, \dots, X_n)$  uma amostra aleatória da variável aleatória  $X \sim U(0, \theta)$ . Temos que

$$L(\theta; \mathbf{x}) = \prod_{i=1}^n \frac{1}{\theta} I_{[0, \theta]}(x_i) = \frac{1}{\theta^n} I_{[0, \theta]}(x_{(n)}) I_{[0, x_{(n)}]}(x_{(1)}), \quad \theta > 0,$$

em que  $x_{(1)} = \min(x_1, \dots, x_n)$  e  $x_{(n)} = \max(x_1, \dots, x_n)$ . Portanto, pelo critério da fatoração,  $X_{(n)} = \max(X_1, \dots, X_n)$  é uma estatística suficiente para  $\theta$ .

**Exemplo 2.1.5. Amostra aleatória de  $N(\mu, \sigma^2)$ ,  $\sigma^2$  conhecido.** Seja  $\mathbf{X} = (X_1, \dots, X_n)$  uma amostra aleatória da distribuição  $N(\mu, \sigma^2)$ . Suponhamos que  $\sigma^2$  seja conhecido (fixado). Temos que

$$\begin{aligned} L(\mu; \mathbf{x}) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \\ &= \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}} \\ &= \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 - \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i}, \end{aligned} \quad (2.4)$$

$\mu \in \mathfrak{R}$ . Portanto, pelo critério da fatoração,  $T(\mathbf{X}) = \sum_{i=1}^n X_i$  é uma estatística suficiente para  $\mu$ .

**Exemplo 2.1.6. Amostra aleatória de  $N(\mu, \sigma^2)$ .** Suponhamos que ambos os parâmetros sejam desconhecidos. Temos, então, que  $\theta = (\mu, \sigma^2)$ . Nesse caso, a função de verossimilhança  $L(\theta; \mathbf{x})$  tem a forma (2.4) com  $\mu \in \mathfrak{R}$  e  $\sigma^2 > 0$ . Tomando  $h(\mathbf{x}) = 1/(\sqrt{2\pi})^n$  e

$$g_{\theta}(t_1(\mathbf{x}), t_2(\mathbf{x})) = \frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i - n \frac{\mu^2}{2\sigma^2}},$$

o critério da fatoração garante que a estatística  $\mathbf{T} = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$  é (conjuntamente) suficiente para  $(\mu, \sigma^2)$ .

**Definição 2.1.2.** Dizemos que duas estatísticas  $T_1$  e  $T_2$  são equivalentes se existir uma relação 1:1 entre elas.

Em outras palavras,  $T_1$  e  $T_2$  são equivalentes se  $T_1$  puder ser obtida a partir de  $T_2$  e vice-versa. Nesse caso, temos que, se  $T_1$  é suficiente para  $\theta$ , então  $T_2$  também é suficiente para  $\theta$ .

**Exemplo 2.1.7. Amostra aleatória de  $N(\mu, \sigma^2)$ ,  $\sigma^2$  conhecido.** Vimos que  $T_1 = \sum_{i=1}^n X_i$  é suficiente para  $\mu$ . Como  $T_1$  é equivalente a  $T_2 = \sum_{i=1}^n X_i/n = \bar{X}$ , temos que  $T_2 = \bar{X}$  também é suficiente para  $\mu$ .

**Exemplo 2.1.8. Amostra aleatória de  $N(\mu, \sigma^2)$ .** Não é difícil verificar que  $\mathbf{T}_1 = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$  e  $\mathbf{T}_2 = (\bar{X}, S^2)$  são equivalentes. Como  $\mathbf{T}_1$  é suficiente para  $\theta = (\mu, \sigma^2)$  (Exemplo 2.1.6), temos que  $\mathbf{T}_2$  também é suficiente para  $\theta$ .

Na família exponencial unidimensional é possível obter estatísticas suficientes unidimensionais. De fato, seja  $\mathbf{X} = (X_1, \dots, X_n)$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$ , com f.d.p. ou f.p. na família exponencial unidimensional (1.1). Então, a distribuição conjunta de  $\mathbf{X}$  é dada por

$$f(\mathbf{x}|\theta) = e^{c^*(\theta)T^*(\mathbf{x}) + d^*(\theta) + S^*(\mathbf{x})}, \quad \mathbf{x} \in A^n,$$

em que  $T^*(\mathbf{x}) = \sum_{i=1}^n T(x_i)$ ,  $c^*(\theta) = c(\theta)$ ,  $d^*(\theta) = nd(\theta)$  e  $S^*(\mathbf{x}) = \sum_{i=1}^n S(x_i)$ , que é da família exponencial unidimensional. Note-se que, tomando

$$h(\mathbf{x}) = e^{\sum_{i=1}^n S(x_i)} \quad \text{e} \quad g_\theta(T(\mathbf{x})) = e^{c(\theta) \sum_{i=1}^n T(x_i) + nd(\theta)},$$

é fácil ver que, pelo critério da fatoração,  $T(\mathbf{X}) = \sum_{i=1}^n T(X_i)$  é uma estatística suficiente para  $\theta$ .

É fácil verificar que amostras aleatórias de famílias exponenciais de dimensão  $k$  também têm distribuições que são

membros da família exponencial com a mesma dimensão. De fato, se  $\mathbf{X} = (X_1, \dots, X_n)$  é uma amostra aleatória de uma variável aleatória com f.d.p. ou f.p. na forma (1.2), temos que a f.d.p. ou f.p. conjunta de  $\mathbf{X}$  é dada por

$$f(x_1, \dots, x_n | \theta) = \exp \left\{ \sum_{j=1}^k c_j^*(\theta) T_j^*(\mathbf{x}) + d^*(\theta) + S^*(\mathbf{x}) \right\},$$

onde  $T_j^*(\mathbf{x}) = \sum_{i=1}^n T_j(x_i)$ ,  $c_j^*(\theta) = c_j(\theta)$ ,  $S^*(\mathbf{x}) = \sum_{i=1}^n S(x_i)$ ,  $d^*(\theta) = nd(\theta)$ . Neste caso,  $(T_1^*, \dots, T_k^*)$  é conjuntamente suficiente para  $\theta$ .

**Exemplo 2.1.9. Amostra aleatória de  $N(\mu, \sigma^2)$ .** Aqui  $\theta = (\mu, \sigma^2)$  e

$$\begin{aligned} f(x|\theta) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ &= \exp \left\{ -\frac{1}{2\sigma^2} x^2 + \frac{\mu}{\sigma^2} x - \frac{\mu^2}{2\sigma^2} - \frac{1}{2} \log \sigma^2 - \log \sqrt{2\pi} \right\}, \end{aligned}$$

que é da família exponencial bidimensional com

$$\begin{aligned} T_1(x) &= x, \quad T_2(x) = x^2, \quad c_1(\theta) = \frac{\mu}{\sigma^2}, \quad c_2(\theta) = -\frac{1}{2\sigma^2}, \\ d(\theta) &= -\frac{\mu^2}{2\sigma^2} - \frac{1}{2} \log \sigma^2, \quad S(x) = -\log \sqrt{2\pi}, \quad A = \mathfrak{R}. \end{aligned}$$

A distribuição de uma amostra aleatória da densidade acima é também da família exponencial bidimensional com  $T_1(\mathbf{X}) = \sum_{i=1}^n X_i$  e  $T_2(\mathbf{X}) = \sum_{i=1}^n X_i^2$ , e  $(T_1, T_2)$  é uma estatística (conjuntamente) suficiente para  $(\mu, \sigma^2)$ .

## 2.2 Propriedades de estimadores

Para a estimação de um parâmetro desconhecido é necessária a escolha de um estimador adequado. Para tal, estudam-

se as propriedades dos estimadores e estabelecem-se critérios de otimalidade.

### 2.2.1 Estimador não viciado e erro quadrático médio

Qualquer estimador é uma variável aleatória e tem, portanto, uma distribuição de probabilidade. Seu valor observado depende da particular amostra extraída da população em estudo. Claramente, não é possível antever se a estimativa produzida pela amostra será próxima ou não do verdadeiro valor do parâmetro que se objetiva estimar. Entretanto, se a distribuição do estimador puder ser determinada, ou se, ao menos, algumas características dessa distribuição puderem ser obtidas, pode ser viável verificar se o estimador possui algumas boas propriedades. Intuitivamente, é desejável que um estimador tenha distribuição centrada em  $\theta$ , não havendo uma tendência a superestimar ou subestimar o parâmetro desconhecido. Em outras palavras, é desejável que o estimador seja não viciado.

**Definição 2.2.1.** *O viés de um estimador  $\hat{\theta}$  do parâmetro  $\theta$  é dado por*

$$B[\hat{\theta}] = E[\hat{\theta}] - \theta, \quad \theta \in \Theta$$

**Definição 2.2.2.** *Dizemos que um estimador  $\hat{\theta}$  é não viciado para  $\theta$  se*

$$E[\hat{\theta}] = \theta, \quad \text{para todo } \theta \in \Theta,$$

*ou seja  $B[\hat{\theta}] = 0$ , para todo  $\theta \in \Theta$ .*

Há situações em que o estimador possui um viés que decresce à medida que o tamanho da amostra cresce. Ou seja, espera-se que, em amostras grandes, o viés de  $\hat{\theta}$  seja muito pequeno. Formalmente, temos a seguinte definição.

**Definição 2.2.3.** Dizemos que um estimador  $\hat{\theta}$  é assintoticamente não viciado para  $\theta$  se

$$\lim_{n \rightarrow \infty} B[\hat{\theta}] = 0, \quad \text{para todo } \theta \in \Theta.$$

**Exemplo 2.2.1. Amostra aleatória de população com média  $\theta$  e variância  $\sigma^2$ .** Seja  $\mathbf{X} = (X_1, \dots, X_n)$  uma amostra aleatória da variável aleatória  $X$  com  $E[X] = \mu$  e  $\text{Var}[X] = \sigma^2$ . Temos que

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \mu$$

e

$$\text{Var}[\bar{X}] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{\sigma^2}{n}.$$

Portanto,  $\bar{X}$  é um estimador não viciado de  $\mu$ . Com relação à variância amostral,  $\hat{\sigma}^2 = \sum_{i=1}^n (X_i - \bar{X})^2/n$ , é possível mostrar que

$$E[\hat{\sigma}^2] = \frac{(n-1)}{n} \sigma^2 \quad \text{e} \quad B[\hat{\sigma}^2] = -\frac{\sigma^2}{n}.$$

Portanto,  $\hat{\sigma}^2$  é um estimador viciado de  $\sigma^2$ , mas é assintoticamente não viciado, ou seja, à medida que o tamanho da amostra aumenta, seu vício diminui. O viés de  $\hat{\sigma}^2$  pode ser corrigido pela multiplicação pelo fator  $n/(n-1)$ , o que resulta no estimador

$$S^2 = \frac{n}{n-1} \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

que é um estimador não viciado para  $\sigma^2$ . É por esta razão que, usualmente, estima-se a variância populacional utilizando-se a variância amostral obtida com  $n-1$  e não com  $n$  no denominador.

Estimadores no viciados podem ser comparados atrs de suas varincias, que so medidas de variabilidade das estimativas em torno do parmetro a ser estimado em amostras repetidas. No entanto, se a comparao envolve ao menos um estimador viciado,  mais adequado utilizar o erro quadrtico mdio.

**Definio 2.2.4.** *O erro quadrtico mdio (EQM) de um estimador  $\hat{\theta}$  do parmetro  $\theta$   dado por*

$$\text{EQM}[\hat{\theta}] = \text{E}[(\hat{\theta} - \theta)^2], \quad \theta \in \Theta.$$

Note-se que

$$\begin{aligned} \text{EQM}[\hat{\theta}] &= \text{E}[\{(\hat{\theta} - \text{E}[(\hat{\theta})]) + (\text{E}[\hat{\theta}] - \theta)\}^2] \\ &= \text{E}[(\hat{\theta} - \text{E}[(\hat{\theta})])^2] + 2\text{E}[(\hat{\theta} - \text{E}[(\hat{\theta})])(\text{E}[\hat{\theta}] - \theta)] \\ &\quad + (\text{E}[\hat{\theta}] - \theta)^2 \\ &= \text{E}[(\hat{\theta} - \text{E}[(\hat{\theta})])^2] + (\text{E}[\hat{\theta}] - \theta)^2 \\ &= \text{Var}[\hat{\theta}] + \text{B}[\hat{\theta}]^2. \end{aligned}$$

Logo, se  $\hat{\theta}$   um estimador no viciado para  $\theta$ , ou seja, se  $\text{B}[\hat{\theta}] = 0$ , o erro quadrtico mdio de  $\hat{\theta}$  se reduz  sua varincia.

**Exemplo 2.2.2. Amostra aleatria de  $N(\mu, \sigma^2)$ .** Conforme visto no Exemplo 2.2.1,  $\hat{\sigma}^2$   um estimador viciado de  $\sigma^2$  enquanto que  $S^2$   no viciado. Por outro lado, temos que

$$\text{EQM}[S^2] = \text{Var}[S^2] = \frac{2\sigma^4}{n-1}$$

e

$$\text{EQM}[\hat{\sigma}^2] = \frac{2\sigma^4}{n-1} \left[ 1 - \frac{3n-1}{2n^2} \right].$$

Note-se que  $\hat{\sigma}^2$ , apesar de ser um estimador viciado, apresenta EQM menor que o estimador  $S^2$ .

### 2.2.2 Estimadores eficientes

Como mencionado anteriormente, a comparação entre estimadores não viciados pode ser feita através de suas variâncias. A eficiência de um estimador pode ser medida pelo quociente entre a menor variância que pode ser atingida por qualquer estimador não viciado e a sua própria variância.

**Definição 2.2.5.** Chamamos de eficiência de um estimador  $\hat{\theta}$ , não viciado para o parâmetro  $\theta$ , o quociente

$$\mathcal{E}[\hat{\theta}] = \frac{\text{LI}(\theta)}{\text{Var}[\hat{\theta}]},$$

em que  $\text{LI}(\theta)$  é o limite inferior da variância dos estimadores não viciados de  $\theta$ . Se  $\mathcal{E}[\hat{\theta}] = 1$ ,  $\hat{\theta}$  é dito ser eficiente.

Como veremos no Teorema 2.2.1,

$$\text{LI}(\theta) = \frac{1}{n\text{E} \left[ \left( \frac{\partial \log f(X|\theta)}{\partial \theta} \right)^2 \right]}, \quad (2.5)$$

quando certas condições de regularidade estão satisfeitas. As condições de regularidade são basicamente duas: o suporte  $A = \{x, f(x|\theta) > 0\}$  é independente de  $\theta$  e é possível a troca da ordem das operações de derivação com respeito a  $\theta$  e de integração sob a distribuição de  $X$ .

**Exemplo 2.2.3. Amostra aleatória de  $N(\mu, \sigma^2)$ ,  $\sigma^2$  conhecido.** Temos que

$$\log f(x|\mu) = -\log \sqrt{2\pi} - \frac{1}{2} \log \sigma^2 - \frac{(x - \mu)^2}{2\sigma^2}.$$

Portanto,

$$\frac{\partial \log f(x|\mu)}{\partial \mu} = \frac{(x - \mu)}{\sigma^2}. \quad (2.6)$$

Assim,

$$\mathbb{E} \left[ \left( \frac{\partial \log f(X|\mu)}{\partial \mu} \right)^2 \right] = \mathbb{E} \left[ \frac{(X - \mu)^2}{\sigma^4} \right] = \frac{1}{\sigma^4} \text{Var}[X] = \frac{1}{\sigma^2}.$$

Logo, conclui-se de (2.5) que

$$\text{LI}(\mu) = \frac{\sigma^2}{n},$$

que coincide com  $\text{Var}[\bar{X}]$ , e, portanto,  $\bar{X}$  é um estimador eficiente para  $\mu$ . De (2.6) temos também que

$$\mathbb{E} \left[ \frac{\partial \log f(X|\mu)}{\partial \mu} \right] = \frac{1}{\sigma^2} \mathbb{E}[X - \mu] = 0. \quad (2.7)$$

**Definição 2.2.6.** *A quantidade*

$$\frac{\partial \log f(X|\theta)}{\partial \theta}$$

*é chamada função escore.*

O resultado (2.7), ou seja, que

$$\mathbb{E} \left[ \frac{\partial \log f(X|\theta)}{\partial \theta} \right] = 0, \quad (2.8)$$

vale em geral quando as condições de regularidade estão satisfeitas. Portanto, o valor esperado da função escore é sempre nulo.

**Definição 2.2.7.** *A quantidade*

$$I_{\text{F}}(\theta) = \mathbb{E} \left[ \left( \frac{\partial \log f(X|\theta)}{\partial \theta} \right)^2 \right],$$

*é denominada informação de Fisher de  $\theta$ .*

Como consequência de (2.8) temos que

$$I_{\text{F}}(\theta) = \text{Var} \left[ \frac{\partial \log f(X|\theta)}{\partial \theta} \right],$$



pois  $\text{Var}[X] = E[X^2]$  para uma variável aleatória  $X$  qualquer com  $E[X] = 0$ . Um resultado importante estabelece que

$$E \left[ \left( \frac{\partial \log f(X|\theta)}{\partial \theta} \right)^2 \right] = -E \left[ \frac{\partial^2 \log f(X|\theta)}{\partial \theta^2} \right].$$

Uma outra propriedade relevante estabelece que, para uma amostra aleatória  $(X_1, \dots, X_n)$  da variável aleatória  $X$  com f.d.p ou f.p.  $f(x|\theta)$  e informação de Fisher  $I_F(\theta)$ , a informação total de Fisher de  $\theta$  correspondente à amostra observada é a soma da informação de Fisher das  $n$  observações da amostra. De fato, sendo

$$L(\theta; \mathbf{x}) = f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta)$$

a f.d.p. conjunta de  $(X_1, \dots, X_n)$ , temos que

$$\begin{aligned} E \left[ \left( \frac{\partial \log L(\theta; \mathbf{X})}{\partial \theta} \right)^2 \right] &= -E \left[ \frac{\partial^2 \log L(\theta; \mathbf{X})}{\partial \theta^2} \right] \\ &= -E \left[ \sum_{i=1}^n \frac{\partial^2 \log f(X_i|\theta)}{\partial \theta^2} \right] = \sum_{i=1}^n E \left[ -\frac{\partial^2 \log f(X_i|\theta)}{\partial \theta^2} \right] = nI_F(\theta), \end{aligned}$$

pois  $X_i$ , para  $i = 1, \dots, n$ , são independentes e identicamente distribuídas com a mesma distribuição que  $X$ .

**Teorema 2.2.1. Desigualdade da Informação.** *Quando as condições de regularidade estão satisfeitas, a variância de qualquer estimador não viciado  $\hat{\theta}$  do parâmetro  $\theta$  satisfaz a desigualdade*

$$\text{Var}[\hat{\theta}] \geq \frac{1}{nI_F(\theta)}.$$

A desigualdade da informação, inicialmente chamada de Cramér-Rao, não é um método de construção de estimadores, mas apenas possibilita verificar se determinado estimador é ou

não eficiente. É então importante que sejam estabelecidos métodos para construção de estimadores que tenham boas propriedades.

### 2.2.3 Estimadores Baseados em Estatísticas Suficientes

Seja  $\mathbf{X} = (X_1, \dots, X_n)$  uma amostra aleatória da variável aleatória  $X$  com f.d.p. ou f.p.  $f(x|\theta)$ . Seja  $T = T(\mathbf{X})$  uma estatística suficiente para  $\theta$  e  $S = S(\mathbf{X})$  um estimador de  $\theta$ . Como  $T$  é uma estatística suficiente, a distribuição condicional de  $S$  dado  $T$  não depende de  $\theta$ . Então,

$$\hat{\theta} = E[S|T] \quad (2.9)$$

também é um estimador de  $\theta$  e depende da amostra somente através de  $T$ . Temos ainda que, se  $S$  é um estimador não viciado de  $\theta$ , então  $\hat{\theta}$  é um estimador não viciado de  $\theta$ . Em outras palavras, a partir de qualquer estimador não viciado  $S$  de  $\theta$ , pode-se encontrar um estimador não viciado que seja função apenas da estatística suficiente  $T$  usando (2.9). Um resultado importante, conhecido como Teorema de Rao-Blackwell, estabelece que, se  $S$  é um estimador não viciado de  $\theta$ , então,

$$\text{Var}[\hat{\theta}] \leq \text{Var}[S], \quad \text{para todo } \theta \in \Theta.$$

Portanto, o estimador  $\hat{\theta}$ , que é baseado na estatística suficiente  $T$ , é não viciado e apresenta variância menor (ou igual) que a variância do estimador não viciado  $S$ . Desse modo, qualquer estimador  $S$  que não seja função de uma estatística suficiente pode ser melhorado pelo procedimento (2.9).

**Exemplo 2.2.4. Amostra aleatória de Poisson( $\theta$ ).** Suponha que o interesse seja estimar  $P[X = 0] = \tau = e^{-\theta}$ . Temos que

a estatística  $T = \sum_{i=1}^n X_i$  é suficiente para  $\theta$  (Exemplo 2.1.2). Consideremos

$$S = \begin{cases} 1, & \text{se } X_1 = 0, \\ 0, & \text{caso contrário.} \end{cases}$$

Temos que  $E[S] = P[X_1 = 0] = \tau$ , ou seja,  $S$  é um estimador não viciado de  $\tau$ . Notemos que, para  $t = 0, 1, 2, \dots$ ,

$$\begin{aligned} E[S|T = t] &= P[X_1 = 0|T = t] = \frac{P[\sum_{i=2}^n X_i = t]P[X_1 = 0]}{P[\sum_{i=1}^n X_i = t]} \\ &= \frac{e^{-(n-1)\theta}((n-1)\theta)^t}{t!} e^{-\theta} \frac{t!}{e^{-n\theta}(n\theta)^t} = \left(\frac{n-1}{n}\right)^t. \end{aligned}$$

Portanto,

$$\hat{\tau} = \left(\frac{n-1}{n}\right)^{\sum_{i=1}^n X_i}$$

é um estimador não viciado de  $\tau$  e é melhor que o estimador  $S$ , pois apresenta EQM menor.

O Teorema de Rao-Blackwell não possibilita garantir que um estimador obtido pelo mecanismo (2.9), sendo  $S$  um estimador não viciado, seja o de menor variância entre todos os estimadores não viciados. O chamado Teorema de Lehmann-Scheffé garante que, se  $T$  for uma estatística completa, além de suficiente, o estimador  $\hat{\theta}$  em (2.9) é o único estimador não viciado de variância uniformemente mínima (ENNVUM) de  $\theta$ .<sup>1</sup> A definição de estatística completa não será fornecida neste texto. É possível demonstrar que, na família exponencial de dimensão  $k$ , como dada na Definição 1.2.3, a estatística  $(T_1(\mathbf{X}), \dots, T_k(\mathbf{X}))$  é completa (já sabemos que é suficiente), desde que o espaço paramétrico contenha retângulos de dimensão  $k$ .

<sup>1</sup> O termo “uniformemente” indica que a mínima variância vale qualquer que seja o valor de  $\theta \in \Theta$ .

**Exemplo 2.2.5. Amostra aleatória de Poisson( $\theta$ ).** Por propriedades da família exponencial, temos que  $T = \sum_{i=1}^n X_i$  é uma estatística suficiente e completa. Como  $\bar{X}$  é o estimador não viesado de  $\theta$  e é função de  $T$ , é ENVVUM de  $\theta$ .

## 2.2.4 Estimadores consistentes

Estimadores consistentes são aqueles que, à medida que o tamanho da amostra aumenta, aproximam-se do parâmetro que está sendo estimado. Consistência está ligada ao conceito de convergência em probabilidade; veja James [6].

**Definição 2.2.8.** *Seja  $\mathbf{X} = (X_1, \dots, X_n)$  uma amostra aleatória da distribuição da variável aleatória  $X$  que depende do parâmetro  $\theta$ . Dizemos que o estimador  $\hat{\theta}$  é consistente para  $\theta$  se*

$$\lim_{n \rightarrow \infty} P[|\hat{\theta} - \theta| > \epsilon] = 0, \quad \text{para qualquer } \epsilon > 0.$$

**Exemplo 2.2.6. Amostra aleatória de população com média  $\theta$  e variância  $\sigma^2$ .** Da desigualdade de Chebyshev (veja James [6]) temos que

$$P[|\bar{X} - \theta| > \epsilon] \leq \frac{\sigma^2}{n\epsilon^2},$$

o que leva a

$$\lim_{n \rightarrow \infty} P[|\bar{X} - \theta| > \epsilon] = 0,$$

e, portanto,  $\bar{X}$  é um estimador consistente para  $\theta$ .

## 2.3 Estimadores de máxima verossimilhança

Nesta seção apresentamos um dos métodos de obtenção de estimadores mais utilizados, o método de máxima verossimilhança. Como vimos na Seção 2.1, a função de verossimilhança

dada em (2.2) é a f.d.p. ou f.p. de  $\mathbf{X} = (X_1, \dots, X_n)$  avaliada na amostra observada  $\mathbf{x} = (x_1, \dots, x_n)$  vista como função do parâmetro desconhecido  $\theta \in \Theta$ . Assim, faz sentido tomar como estimativa de  $\theta$  o valor em  $\Theta$  que maximiza a função de verossimilhança. No caso discreto, esta estimativa pode ser interpretada como o valor de  $\theta$  que maximiza a probabilidade de se observar a amostra que foi selecionada.

**Definição 2.3.1.** *Seja  $L(\theta; \mathbf{x})$  a função de verossimilhança correspondente à amostra observada  $\mathbf{x}$ . A estimativa de máxima verossimilhança de  $\theta$  é o valor  $\hat{\theta} \in \Theta$  que maximiza  $L(\theta; \mathbf{x})$ ;  $\hat{\theta}$  avaliado em  $\mathbf{X}$  é chamado de estimador de máxima verossimilhança de  $\theta$ .*

O logaritmo natural da função de verossimilhança de  $\theta$  é denotado por  $l(\theta; \mathbf{x}) = \log L(\theta; \mathbf{x})$ . Como o logaritmo é uma função monótona crescente, qualquer valor de  $\theta$  que maximiza  $l(\theta; \mathbf{x})$  também maximiza a função de verossimilhança  $L(\theta; \mathbf{x})$ . Além disso, quando  $\Theta$  é um intervalo aberto da reta e  $l(\theta; \mathbf{x})$  é derivável, uma estimativa de máxima verossimilhança pode, em muitos casos, ser encontrada como raiz da equação de verossimilhança

$$l'(\theta; \mathbf{x}) = \frac{\partial l(\theta; \mathbf{x})}{\partial \theta} = 0. \quad (2.10)$$

Para se concluir que uma solução desta equação é um ponto de máximo é necessário verificar se

$$l''(\hat{\theta}; \mathbf{x}) = \left. \frac{\partial^2 \log L(\theta; \mathbf{x})}{\partial \theta^2} \right|_{\theta=\hat{\theta}} < 0. \quad (2.11)$$

Em situações em que  $\Theta$  é discreto ou em que o máximo de  $l(\theta; \mathbf{x})$  ocorre na fronteira do espaço paramétrico  $\Theta$ , o estimador de máxima verossimilhança pode ser obtido por inspeção da função de verossimilhança.

**Exemplo 2.3.1. Amostra aleatória de  $N(\mu, \sigma^2)$ ,  $\sigma^2$  conhecido.** A função de verossimilhança é dada por

$$L(\mu; \mathbf{x}) = \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}, \quad -\infty < \mu < \infty.$$

Como

$$l(\mu; \mathbf{x}) = -n \log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2,$$

segue que a equação de verossimilhança é dada por

$$\sum_{i=1}^n (x_i - \hat{\mu}) = 0.$$

Logo, o estimador de máxima verossimilhança de  $\mu$  é dado por

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

Não é difícil verificar que (2.11) está satisfeita.

Portanto,  $\bar{X}$ , além de ser eficiente (Exemplo 2.2.3) e função da estatística suficiente, é também estimador de máxima verossimilhança.

**Exemplo 2.3.2. Amostra aleatória de Bernoulli( $\theta$ ).** A função de verossimilhança de  $\theta$  é dada por

$$L(\theta; \mathbf{x}) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}, \quad 0 < \theta < 1.$$

Assim,

$$l(\theta; \mathbf{x}) = \sum_{i=1}^n x_i \log \theta + \left( n - \sum_{i=1}^n x_i \right) \log(1 - \theta).$$

Logo, a equação de verossimilhança de  $\theta$  é dada por

$$\frac{\sum_{i=1}^n x_i}{\hat{\theta}} - \frac{(n - \sum_{i=1}^n x_i)}{1 - \hat{\theta}} = 0.$$

Portanto, o estimador de máxima verossimilhança de  $\theta$  é  $\hat{\theta} = \bar{X}$ , pois, neste caso, (2.11) também está verificada.

O exemplo a seguir ilustra uma situação em que a equação (2.10) não pode ser utilizada.

**Exemplo 2.3.3. Amostra aleatória de  $U(0, \theta)$ .** Como visto no Exemplo 2.1.4, podemos escrever a função de verossimilhança como

$$L(\theta; \mathbf{x}) = \frac{1}{\theta^n} I_{[0, \theta]}(x_{(n)}) I_{[0, x_{(n)}]}(x_{(1)}), \quad \theta > 0.$$

Como a função de verossimilhança é nula para  $\theta < x_{(n)}$  e é decrescente para  $\theta \geq x_{(n)}$ , o máximo de  $L(\theta; \mathbf{x})$  é dado por  $x_{(n)}$ . Logo, o estimador de máxima verossimilhança de  $\theta$  é  $\hat{\theta} = X_{(n)}$ , que é uma estatística suficiente para  $\theta$ . Nesse caso o estimador de máxima verossimilhança de  $\theta$  é viciado.

Nos exemplos apresentados acima, a solução da equação de verossimilhança pode ser obtida explicitamente. Em alguns casos, principalmente quando a verossimilhança está associada a modelos mais complexos, a função de verossimilhança não apresenta solução analítica explícita. Em tais casos, os estimadores de máxima verossimilhança podem ser obtidos por meio de métodos numéricos (por exemplo, método de Newton-Raphson ou método do escore).

### 2.3.1 Propriedades dos Estimadores de Máxima Verossimilhança

O teorema a seguir apresenta uma propriedade importante dos estimadores de máxima verossimilhança, estabelecendo que o estimador de máxima verossimilhança é função de qualquer estatística suficiente.

**Teorema 2.3.1.** *Seja  $\mathbf{X} = (X_1, \dots, X_n)$  uma amostra aleatria da varivel aleatria  $X$  com f.d.p. ou f.p.  $f(x|\theta)$ . Seja  $T = T(\mathbf{X})$  uma estatística suficiente para  $\theta$ . Ento o estimador de mxima verossimilhana  $\hat{\theta}$  (se existir) é funo de  $T$ .*

**Prova.** Pelo critrio da fatoraco, se  $T$  é suficiente para  $\theta$ , ento,

$$L(\theta; \mathbf{x}) = h(\mathbf{x})g_\theta(T(\mathbf{x})).$$

Como  $h(\mathbf{x})$  é constante em  $\theta$ , maximizar  $L(\theta; \mathbf{x})$  com relao a  $\theta$  é equivalente a maximizar  $g_\theta(T(\mathbf{x}))$  com relao a  $\theta$ . Como  $g_\theta(T(\mathbf{x}))$  depende de  $\mathbf{x}$  somente atravs de  $T$ , temos que  $\hat{\theta}$  é funo de  $T$ .

Uma outra propriedade interessante dos estimadores de mxima verossimilhana é a de invarincia, ou seja, se  $\hat{\theta}$  é um estimador de mxima verossimilhana de  $\theta$ , ento  $g(\hat{\theta})$  é um estimador de mxima verossimilhana de  $g(\theta)$ .

**Exemplo 2.3.6. Amostra aleatria de Bernoulli( $\theta$ ).** Suponha que o interesse seja estimar  $g(\theta) = \text{Var}[X] = \theta(1 - \theta)$ . Pela propriedade de invarincia, temos que o estimador de mxima verossimilhana de  $g(\theta)$  é

$$g(\hat{\theta}) = \bar{X}(1 - \bar{X}).$$

É possvel mostrar que  $g(\hat{\theta})$  é um estimador viciado para  $g(\theta)$ . Por outro lado, mostra-se tambm que

$$E[g(\hat{\theta})] - g(\theta) = \frac{1}{n}\theta(1 - \theta),$$

que decresce  medida que  $n$  aumenta.

**Exemplo 2.3.7. Amostra aleatria de  $N(\mu, \sigma^2)$ ,  $\sigma^2$  conhecido.** Vimos que  $\hat{\mu} = \bar{X}$  é o estimador de mxima verossimi-



lhança de  $\mu$ . Suponha que o interesse seja estimar

$$g(\mu) = P_\mu[X \leq 0] = \Phi(-\mu).$$

Pela propriedade de invariância, temos que

$$g(\hat{\mu}) = \Phi(-\bar{X})$$

é o estimador de máxima verossimilhança de  $g(\mu)$ .

**Exemplo 2.3.8. Amostra aleatória de  $\text{Exp}(\theta)$ .** Seja  $(X_1, \dots, X_n)$  uma amostra aleatória da distribuição da variável aleatória  $X$  com distribuição exponencial de parâmetro  $\theta$ . Neste caso,  $\hat{\theta} = 1/\bar{X}$  é o estimador de máxima verossimilhança de  $\theta$ . Suponha que o interesse seja estimar

$$g(\theta) = P_\theta[X > 1] = e^{-\theta}.$$

Pelo princípio da invariância, o estimador de máxima verossimilhança  $g(\theta)$  é

$$g(\hat{\theta}) = e^{-1/\bar{X}}.$$

Nos três exemplos acima, o estimador de máxima verossimilhança é uma função complicada da amostra. Certamente, não é uma tarefa fácil encontrar a distribuição do estimador  $\Phi(-\bar{X})$ , por exemplo. Contudo, se o tamanho da amostra for grande, o estimador de máxima verossimilhança apresentará uma distribuição aproximadamente normal, como veremos adiante. Além disso, veremos que o estimador de máxima verossimilhança é aproximadamente eficiente em grandes amostras.

Se condições de regularidade estão satisfeitas, temos que

$$\sqrt{n}(\hat{\theta} - \theta) \stackrel{a}{\sim} N\left(0, \frac{1}{I_F(\theta)}\right), \quad (2.12)$$

e

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \stackrel{a}{\sim} N\left(0, \frac{(g'(\theta))^2}{I_F(\theta)}\right), \quad (2.13)$$

em que  $\stackrel{a}{\sim}$  indica distribuição assintótica. Em outras palavras, em amostras grandes, os estimadores de máxima verossimilhança de  $\theta$  e de  $g(\theta)$  têm distribuição aproximadamente normal, são aproximadamente não viciados, e têm variâncias próximas dos correspondentes limites inferiores das variâncias dos estimadores não viciados. Portanto, em grandes amostras, os estimadores de máxima verossimilhança são aproximadamente eficientes.

**Exemplo 2.3.9. Amostra aleatória de Poisson( $\theta$ ).** O estimador de máxima verossimilhança de  $\theta$  é  $\hat{\theta} = \bar{X}$  (verifique!). Pela propriedade de invariância, temos que o estimador de máxima verossimilhança de  $P_\theta[X = 0] = e^{-\theta}$  é dado por

$$g(\hat{\theta}) = e^{-\bar{X}}.$$

De (2.13), temos que

$$\sqrt{n}(g(\hat{\theta}) - e^{-\theta}) \sim N(0, \theta e^{-2\theta}),$$

aproximadamente, se  $n$  for grande.

### 2.3.2 O Caso Multiparamétrico

Nas seções anteriores discutimos a obtenção dos estimadores de máxima verossimilhança e estudamos suas propriedades no caso em que a função de verossimilhança depende apenas de um parâmetro. Nesta seção vamos considerar situações em que  $\theta = (\theta_1, \dots, \theta_k)$ , ou seja, a verossimilhança depende de dois ou mais parâmetros. O espaço paramétrico será denotado por  $\Theta$ . Nos casos em que condições de regularidade estão satisfeitas, os

estimadores de máxima verossimilhança de  $\theta_1, \dots, \theta_k$  podem ser obtidos como solução das equações

$$\frac{\partial l(\theta; \mathbf{x})}{\partial \theta_i} = 0, \quad i = 1, \dots, k.$$

Nos casos em que o suporte da distribuição de  $X$  depende de  $\theta$  ou o máximo ocorre na fronteira de  $\Theta$ , o estimador de máxima verossimilhança é, em geral, obtido inspecionando o gráfico da função de verossimilhança como no caso uniparamétrico. Nos casos em que a função de verossimilhança depende de dois parâmetros,  $\theta_1$  e  $\theta_2$ , utilizando a equação

$$\frac{\partial l(\theta_1, \theta_2; \mathbf{x})}{\partial \theta_1} = 0,$$

obtemos uma solução para  $\theta_1$  como função de  $\theta_2$ , que podemos denotar por  $\hat{\theta}_1(\theta_2)$ . Substituindo a solução para  $\theta_1$  na verossimilhança conjunta, temos agora uma função apenas de  $\theta_2$ , ou seja,

$$g(\theta_2; \mathbf{x}) = l(\hat{\theta}_1(\theta_2), \theta_2; \mathbf{x}),$$

que é o logaritmo da chamada função de verossimilhança perfilada de  $\theta_2$ . Esta pode ser usada para obter o estimador de máxima verossimilhança de  $\theta_2$ . A maximização de  $g(\theta_2; \mathbf{x})$  pode, então, ser feita de maneira usual, ou seja, através de derivação, quando possível.

**Exemplo 2.3.10. Amostra aleatória de  $N(\mu, \sigma^2)$ .** Aqui,

$$l(\mu, \sigma^2; \mathbf{x}) = -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}.$$

Assim, a equação

$$\frac{\partial l(\mu, \sigma^2; \mathbf{x})}{\partial \mu} = 2 \sum_{i=1}^n \frac{(x_i - \hat{\mu})}{2\sigma^2} = 0$$

leva ao estimador  $\hat{\mu} = \bar{X}$ . Logo, o logaritmo da função de verossimilhança perfilada de  $\sigma^2$  é dado por

$$g(\sigma^2; \mathbf{x}) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Portanto, o estimador de máxima verossimilhança de  $\sigma^2$  é obtido como solução da equação

$$\frac{\partial g(\sigma^2; \mathbf{x})}{\partial \sigma^2} = -\frac{n}{2\hat{\sigma}^2} + \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{2\hat{\sigma}^4} = 0$$

que leva ao estimador

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

No caso multiparamétrico, as mesmas propriedades, tais como invariância, função da estatística suficiente e outras, continuam valendo.

## 3 Estimação Intervalar

Neste capítulo consideramos o problema de estimação de parâmetros utilizando intervalos de confiança. Os intervalos de confiança são obtidos a partir de variáveis aleatórias especiais chamadas de quantidades pivotais.

### 3.1 Método da quantidade pivotal

A construção de intervalos utilizando quantidades pivotais é considerada a seguir.

**Definição 3.1.1.** *Uma variável aleatória  $Q(\mathbf{X}; \theta)$  é dita ser uma quantidade pivotal para o parâmetro  $\theta$  se sua distribuição for independente de  $\theta$ .*

Note-se que uma quantidade pivotal não é uma estatística, pois depende do parâmetro  $\theta$ .

Seja  $\alpha \in (0, 1)$  e defina  $\gamma = 1 - \alpha$ . Para  $\gamma$  fixado, sejam  $\lambda_1$  e  $\lambda_2$  tais que

$$P[\lambda_1 \leq Q(\mathbf{X}; \theta) \leq \lambda_2] = \gamma. \quad (3.1)$$

Como a distribuição de  $Q(\mathbf{X}; \theta)$  é independente de  $\theta$ ,  $\lambda_1$  e  $\lambda_2$  também não dependem de  $\theta$ . Além disso, se para cada  $\mathbf{X}$  existirem  $t_1(\mathbf{X})$  e  $t_2(\mathbf{X})$  tais que

$$\lambda_1 \leq Q(\mathbf{X}; \theta) \leq \lambda_2 \text{ se e somente se } t_1(\mathbf{X}) \leq \theta \leq t_2(\mathbf{X}),$$

então, de (3.1) segue que

$$P[t_1(\mathbf{X}) \leq \theta \leq t_2(\mathbf{X})] = \gamma.$$

Assim,  $[t_1(\mathbf{X}); t_2(\mathbf{X})]$  é um intervalo aleatório que contém  $\theta$  com probabilidade  $\gamma$ , e  $\gamma$  é denominado coeficiente de confiança.

Nos casos em que a distribuição da quantidade pivotal é discreta, nem sempre é possível determinar  $\lambda_1$  e  $\lambda_2$  tais que (3.1) esteja satisfeita exatamente. Nesses casos, podemos escolher  $\lambda_1$  e  $\lambda_2$  tais que (3.1) seja satisfeita para um coeficiente de confiança maior ou igual a  $\gamma$ , o mais próximo possível de  $\gamma$ .

Quando  $n$  é razoavelmente grande, intervalos de confiança aproximados podem ser obtidos através da distribuição assintótica do estimador de máxima verossimilhança, vista na Seção 2.3.1. Um outro ponto a salientar é que, na maioria dos casos, existem infinitos pares  $(\lambda_1, \lambda_2)$  satisfazendo (3.1). Sempre que possível, devemos escolher  $(\lambda_1, \lambda_2)$  que produz o intervalo de menor comprimento. Tal procedimento é facilitado em situações em que a distribuição de  $Q(\mathbf{X}; \theta)$  é simétrica, como no caso da distribuição normal.

**Exemplo 3.1.1. Amostra aleatória de  $U(0, \theta)$ .** Vimos no Capítulo 2 que uma estatística suficiente para  $\theta$  é dada por  $Y = X_{(n)}$ . A f.d.p. de  $Y$  é

$$f_Y(y) = \frac{ny^{n-1}}{\theta^n}, \quad y \in [0, \theta].$$

Logo  $X_{(n)}$  não é uma quantidade pivotal já que sua distribuição depende de  $\theta$ . Por outro lado, a distribuição de  $Q(\mathbf{X}; \theta) = X_{(n)}/\theta$  é dada por

$$f_Q(q) = nq^{n-1}I_{[0,1]}(q),$$

que não depende de  $\theta$ , e, portanto,  $Q(\mathbf{X}; \theta)$  é uma quantidade pivotal. Assim, dado  $\gamma = 1 - \alpha$ , podemos encontrar  $\lambda_1$  e  $\lambda_2$  tais

que

$$\int_{\lambda_1}^{\lambda_2} f_Q(q) dq = \gamma = 1 - \alpha.$$

Como existem infinitos pares  $(\lambda_1, \lambda_2)$  satisfazendo esta equação, consideramos o intervalo simétrico, ou seja, consideramos o intervalo satisfazendo

$$\int_0^{\lambda_1} f_Q(q) dq = \frac{\alpha}{2} \quad \text{e} \quad \int_{\lambda_2}^1 f_Q(q) dq = \frac{\alpha}{2},$$

o que leva a

$$\lambda_1 = \left(\frac{\alpha}{2}\right)^{1/n} \quad \text{e} \quad \lambda_2 = \left(1 - \frac{\alpha}{2}\right)^{1/n}.$$

Como

$$\mathbf{P} \left[ \lambda_1 \leq \frac{X_{(n)}}{\theta} \leq \lambda_2 \right] = \mathbf{P} \left[ \frac{X_{(n)}}{\lambda_2} \leq \theta \leq \frac{X_{(n)}}{\lambda_1} \right] = 1 - \alpha,$$

temos que

$$\left[ \frac{X_{(n)}}{(1 - \alpha/2)^{1/n}}; \frac{X_{(n)}}{(\alpha/2)^{1/n}} \right]$$

é um intervalo de confiança para  $\theta$  com coeficiente de confiança  $\gamma = 1 - \alpha$ .

Conforme mencionado anteriormente, o coeficiente de confiança,  $\gamma$ , é a probabilidade fixada de que o intervalo aleatório  $[t_1(\mathbf{X}); t_2(\mathbf{X})]$  contenha o verdadeiro valor do parâmetro. Após a amostra ser observada, o intervalo de confiança é calculado com base nos dados observados e passa a não depender de nenhuma quantidade aleatória; é apenas um intervalo numérico. Desta maneira,  $\gamma$  não pode ser interpretado como a probabilidade do verdadeiro valor de  $\theta$  pertencer ao intervalo de confiança obtido. O que se aplica é a interpretação frequentista, ou seja,

para cada 100 intervalos numricos construdos a partir do intervalo aleatrio, aproximadamente  $100\gamma\%$  deles contero o valor verdadeiro de  $\theta$ . Assim,  $\gamma$  fornece um grau de confiana de que o intervalo de confiana obtido contenha o verdadeiro valor de  $\theta$ .

## 3.2 Intervalos para Populaes Normais

### 3.2.1 O caso de uma nica amostra

Seja  $(X_1, \dots, X_n)$  uma amostra aleatria de tamanho  $n$  da distribuo  $N(\mu, \sigma^2)$ . Assumindo  $\sigma^2$  conhecido, temos que uma quantidade pivotal baseada na estatstica suficiente  $\bar{X}$   dada por

$$Q(\mathbf{X}; \mu) = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

que tem distribuo  $N(0, 1)$ . Portanto, dado o coeficiente de confiana  $\gamma = 1 - \alpha$ , existem  $\lambda_1$  e  $\lambda_2$  tais que

$$P \left[ \lambda_1 \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \lambda_2 \right] = \gamma.$$

Como a distribuo  $N(0, 1)$   simtrica, o intervalo de menor comprimento  o intervalo simtrico, que  obtido tomando  $\lambda_1 = -z_{\alpha/2}$  e  $\lambda_2 = z_{\alpha/2}$ , em que  $P[Z \leq z_{\alpha/2}] = 1 - \alpha/2$  com  $Z \sim N(0, 1)$ . O intervalo de confiana para  $\mu$  de menor comprimento , portanto, dado por

$$\left[ \bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right].$$

Se  $\sigma^2$   desconhecido, temos pelo Teorema 1.5.1.(iii) que

$$Q(\mathbf{X}, \mu) = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$



e, portanto,  $Q(\mathbf{X}, \mu)$  é uma quantidade pivotal. Então, existem  $\lambda_1$  e  $\lambda_2$  tais que

$$P \left[ \lambda_1 \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq \lambda_2 \right] = \gamma.$$

Como a distribuição  $t_{n-1}$  é simétrica, o intervalo de confiança de menor comprimento é obtido tomando  $\lambda_1 = -t_{\alpha/2}$  e  $\lambda_2 = t_{\alpha/2}$ , em que  $t_{\alpha/2}$  é tal que  $P(T \leq t_{\alpha/2}) = 1 - \alpha/2$  com  $T \sim t_{n-1}$ . Assim, o intervalo de menor comprimento é dado por

$$\left[ \bar{X} - t_{\alpha/2} \frac{S}{\sqrt{n}}; \bar{X} + t_{\alpha/2} \frac{S}{\sqrt{n}} \right].$$

Suponha que o interesse seja estimar  $\sigma^2$  sendo  $\mu$  desconhecido. Do Teorema 1.5.1.(ii), temos que

$$Q(\mathbf{X}, \sigma^2) = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

e é, então, uma quantidade pivotal. Portanto, podemos determinar  $\lambda_1$  e  $\lambda_2$  de modo que

$$P \left[ \lambda_1 \leq \frac{(n-1)S^2}{\sigma^2} \leq \lambda_2 \right] = \gamma.$$

Considerando o intervalo simétrico, ou seja,  $\lambda_1 = q_1$  e  $\lambda_2 = q_2$ , em que  $P[\chi_{n-1}^2 \geq q_2] = P[\chi_{n-1}^2 \leq q_1] = \alpha/2$ , obtém-se o intervalo

$$\left[ \frac{(n-1)S^2}{q_2}; \frac{(n-1)S^2}{q_1} \right].$$

### 3.2.2 Duas amostras independentes

Sejam  $\mathbf{X} = (X_1, \dots, X_n)$  e  $\mathbf{Y} = (Y_1, \dots, Y_m)$  amostras aleatórias independentes de  $X \sim N(\mu_1, \sigma^2)$  e  $Y \sim N(\mu_2, \sigma^2)$ , respectivamente. O objetivo é estimar a diferença das médias

populacionais, ou seja,  $\mu_1 - \mu_2$ . Admitindo que a variância  $\sigma^2$ , comum às duas populações, é conhecida, o vetor de parâmetros é  $\theta = (\mu_1, \mu_2)$ . Sabe-se que

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)\right).$$

Assim,

$$Q(\mathbf{X}, \mathbf{Y}, \theta) = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1),$$

e é, portanto, uma quantidade pivotal. Analogamente ao exposto na seção anterior, mostra-se que

$$\left[ \bar{X} - \bar{Y} - z_{\alpha/2}\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}; \bar{X} - \bar{Y} + z_{\alpha/2}\sigma\sqrt{\frac{1}{n} + \frac{1}{m}} \right]$$

é um intervalo de confiança para  $\mu_1 - \mu_2$  com coeficiente de confiança  $\gamma$ .

Se  $\sigma^2$  é desconhecido, temos que uma quantidade pivotal é dada por

$$Q(\mathbf{X}, \mathbf{Y}, \theta) = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}, \quad (3.2)$$

em que

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2},$$

com

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{e} \quad S_y^2 = \frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2.$$

De fato, como

$$\frac{(n-1)S_x^2}{\sigma^2} \sim \chi_{n-1}^2 \quad \text{e} \quad \frac{(m-1)S_y^2}{\sigma^2} \sim \chi_{m-1}^2,$$

e, pela independência de  $S_x^2$  e  $S_y^2$ , temos que

$$\frac{(n+m-2)S_p^2}{\sigma^2} = \frac{(n-1)S_x^2 + (m-1)S_y^2}{\sigma^2} \sim \chi_{n+m-2}^2. \quad (3.3)$$

Então, do Teorema 1.5.1(iii) segue (3.2). Um intervalo de confiança para  $\mu_1 - \mu_2$  com coeficiente de confiança  $\gamma$  é, então, dado por

$$\left[ \bar{X} - \bar{Y} - t_{\alpha/2} S_p \sqrt{\frac{1}{n} + \frac{1}{m}}; \bar{X} - \bar{Y} + t_{\alpha/2} S_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right],$$

em que  $t_{\alpha/2}$  é tal que  $P(T \leq t_{\alpha/2}) = 1 - \alpha/2$  com  $T \sim t_{n+m-2}$ .

Para construir um intervalo de confiança para  $\sigma^2$ , pode-se considerar a quantidade pivotal (3.3).

### 3.3 Intervalos de Confiança Aproximados

Intervalos de confiança aproximados podem ser construídos com base na distribuição assintótica de estimadores de máxima verossimilhança. De (2.12) temos que

$$\sqrt{nI_F(\theta)} (\hat{\theta} - \theta) \stackrel{a}{\sim} N(0, 1).$$

Como  $I_F(\theta)$  pode depender de  $\theta$ , que não é conhecido, substituindo  $I_F(\theta)$  por  $I_F(\hat{\theta})$ , temos que

$$Q(\mathbf{X}, \theta) = \sqrt{nI_F(\hat{\theta})} (\hat{\theta} - \theta) \stackrel{a}{\sim} N(0, 1), \quad (3.4)$$

de modo que  $Q(\mathbf{X}, \theta)$  é uma quantidade aproximadamente pivotal em grandes amostras.

Se o interesse estiver na construção de intervalos de confiança para uma função  $g(\theta)$ , pode-se considerar a variável alea-

tória

$$Q(\mathbf{X}, g(\theta)) = \frac{g(\hat{\theta}) - g(\theta)}{\sqrt{\frac{(g'(\hat{\theta}))^2}{nI_F(\hat{\theta})}}} \stackrel{a}{\sim} N(0, 1), \quad (3.5)$$

que é uma quantidade aproximadamente pivotal para amostras grandes.

**Exemplo 3.3.1. Amostra aleatória de Bernoulli( $\theta$ ).** Como o estimador de máxima verossimilhança de  $\theta$  é  $\hat{\theta} = \bar{X}$  e  $I_F(\theta) = 1/[\theta(1 - \theta)]$ , de (3.4) temos que

$$Q(\mathbf{X}, \theta) = \frac{\bar{X} - \theta}{\sqrt{\frac{\bar{X}(1 - \bar{X})}{n}}} \stackrel{a}{\sim} N(0, 1).$$

Assim, para valores grandes de  $n$ , um intervalo de confiança para  $\theta$  com coeficiente de confiança aproximadamente  $\gamma$  é dado por

$$\left[ \bar{X} - z_{\alpha/2} \sqrt{\frac{\bar{X}(1 - \bar{X})}{n}}; \bar{X} + z_{\alpha/2} \sqrt{\frac{\bar{X}(1 - \bar{X})}{n}} \right].$$

**Exemplo 3.3.2. Amostra aleatória de Exp( $\theta$ ).** Suponha que o interesse seja estimar  $g(\theta) = E[X] = 1/\theta$ . Como  $I_F(\theta) = 1/\theta^2$  e  $\hat{\theta} = 1/\bar{X}$ , segue de (3.5) que

$$Q(\mathbf{X}, \theta) = \frac{\sqrt{n}(\bar{X} - 1/\theta)}{\bar{X}} \stackrel{a}{\sim} N(0, 1),$$

o que conduz ao intervalo de confiança com coeficiente de confiança aproximado  $\gamma = 1 - \alpha$  dado por

$$\left[ \bar{X} - z_{\alpha/2} \frac{\bar{X}}{\sqrt{n}}; \bar{X} - z_{\alpha/2} \frac{\bar{X}}{\sqrt{n}} \right].$$

# Considerações Finais e Notas Bibliográficas

O material apresentado neste livro foca em alguns conceitos centrais da inferência estatística clássica. São abordados os tópicos de estimação pontual e intervalar, enquanto que o tópico de testes de hipóteses, igualmente relevante, é omitido. Espera-se que o estudo deste breve texto motive o leitor a se aprofundar mais e procurar livros e cursos mais abrangentes.

Grande parte do texto é baseada no livro “Introdução à Inferência Estatística” de Bolfarine e Sandoval [1], que é frequentemente utilizado em disciplinas de inferência estatística em nível de graduação no Brasil. Recomendamos ao leitor interessado que leia nesse livro o Capítulo 6, que trata de testes de hipóteses, e as Seções 4.3 e 4.4, que introduzem o enfoque Bayesiano.

Outras referências úteis para aprofundamento do conteúdo aqui apresentado incluem os livros de Casella e Berger [2], DeGroot e Schervish [3], Garthwaite et. al. [4] e Hogg et al. [5].



# Referências

- [1] H. Bolfarine, M.C. Sandoval, *Introdução à Inferência Estatística*. SBM, 2. ed., Rio de Janeiro, 2010.
- [2] G. Casella, R. Berger *Statistical Inference*. Duxbury, 2. ed., 2002.
- [3] M.H. DeGroot, M.J. Schervish, *Probability and Statistics*, Pearson, 4. ed., 2011
- [4] P.H. Garthwaite, I.T. Jolliffe, B. Jones, *Statistical Inference*. Oxford University Press, 2. ed., 2002.
- [5] R.V. Hogg, J. McKean, A.T. Craig, *Introduction to Mathematical Statistics*. Pearson, 7. ed., 2012.
- [6] B.R. James, *Probabilidade: um Curso em Nível Intermediário*. IMPA, 3. ed., Rio de Janeiro, 2008.