

Heleno Bolfarine

Mônica Carneiro Sandoval

# INTRODUÇÃO À INFERÊNCIA ESTATÍSTICA



## CONTEÚDO

PREFÁCIO .....	iv
CAPÍTULO 1. ELEMENTOS BÁSICOS .....	1
1.1. Alguns Modelos Especiais.....	1
1.1.1. O modelo normal .....	1
1.1.2. O modelo exponencial .....	1
1.1.3. O modelo binomial .....	2
1.1.4. O modelo de Poisson .....	3
1.1.5. O modelo uniforme .....	3
1.2. Tipos de Problemas .....	4
1.3. Amostras, Estatísticas e Estimadores .....	5
1.4. Exercícios .....	17
CAPÍTULO 2. ESTIMADORES EFICIENTES E ESTATÍSTICAS SUFICIENTES .....	20
2.1. Estimadores Eficientes .....	20
2.2. Estatísticas Suficientes .....	26
2.3. Estatísticas Conjuntamente Suficientes .....	31
2.4. Famílias Exponenciais .....	33
2.5. Estimadores Baseados em Estatísticas Suficientes .....	39
2.6. Exercícios .....	42
CAPÍTULO 3. MÉTODOS DE ESTIMAÇÃO .....	46
3.1. O Método de Máxima Verossimilhança .....	46
3.2. Propriedades dos Estimadores de Máxima Verossimilhança .....	55
3.2.1. Invariância .....	56
3.2.2. Distribuição em grandes amostras .....	58
3.3. Verossimilhança para Amostras Independentes .....	59
3.4. O Caso Multiparamétrico.....	61
3.5. Família Exponencial e o Método de Máxima Verossimilhança .....	64
3.6. O Método dos Momentos .....	66
3.7. Estimadores Consistentes .....	68
3.8. Exercícios .....	69
CAPÍTULO 4. INTRODUÇÃO À TEORIA DAS DECISÕES. OS PRINCÍPIOS MINIMAX E DE BAYES .....	74
4.1. Os Elementos Básicos .....	49
4.2. O Princípio Minimax .....	52
4.3. O Princípio de Bayes .....	53
4.4. Estimadores de Bayes com Perda Quadrática .....	54

4.5. Exercícios .....	61
CAPÍTULO 5. ESTIMAÇÃO POR INTERVALO .....	96
5.1. Amostras de Populações Normais .....	96
5.2. O Método da Quantidade Pivotal .....	99
5.3. Intervalos para Populações Normais .....	105
5.3.1. O caso de uma única amostra .....	106
5.3.2. Duas amostras independentes .....	107
5.4. Intervalos de Confiança Aproximados .....	109
5.5. Intervalos de Confiança Bayesianos .....	111
5.6. Exercícios .....	114
CAPÍTULO 6. TESTES DE HIPÓTESES .....	118
6.1. Idéias Básicas .....	118
6.2. Formulação Estatística .....	119
6.3. Hipótese Nula Simples contra Alternativa Simples. Testes Mais Poderosos .....	123
6.4. Testes Uniformemente Mais Poderosos .....	130
6.4.1. Hipótese nula simples contra alternativa composta .....	130
6.4.2. Hipóteses compostas .....	133
6.5. Testes da Razão de Verossimilhanças Generalizada .....	134
6.6. Testes Bayesianos .....	149
6.7. Exercícios .....	150
REFERÊNCIAS .....	155
ÍNDICE REMISSIVO .....	156

## PREFÁCIO

O objetivo principal deste texto é propiciar aos estudantes um material básico para um curso introdutório de Inferência Estatística usualmente ministrado em programas de bacharelado em Estatística. Lecionando há vários anos a referida disciplina em cursos de bacharelado e de pós graduação no Departamento de Estatística do Instituto de Matemática e Estatística da Universidade de São Paulo, experimentamos várias alternativas didáticas, mas sempre nos ressentimos da ausência de textos adequados em português e até mesmo em inglês para o nível em questão. E foi pensando em preencher essa lacuna que resolvemos elaborar este trabalho, destinado aos estudantes com conhecimentos básicos de probabilidade e cálculo.

O texto está elaborado para um curso de um semestre com seis horas semanais, duas das quais devem ser reservadas para exercícios. É dividido em seis capítulos, tendo no final de cada um uma série de exercícios.

O Capítulo 1 é dedicado à descrição de alguns modelos comumente utilizados em situações práticas. São apresentados métodos de comparação entre estimadores, com ênfase especial ao método do Erro Quadrático Médio mínimo.

O Capítulo 2 apresenta à obtenção de estimadores eficientes, utilizando a desigualdade da informação, a partir da qual se obtém o limite inferior da variância dos estimadores não viciados. Usando esses resultados em alguns modelos importantes, é possível a obtenção de estimadores ótimos, ou seja, de menor variância. Uma família importante em que tais estimadores são obtidos é a bem conhecida família exponencial de distribuições, apresentada no texto com algum detalhe. A utilização de estatísticas suficientes, no sentido de apresentarem um resumo dos dados sem perda de informação, é também considerada nesse capítulo. Mostra-se também que estimadores que não são funções de estatísticas suficientes podem ser melhorados por meio do conhecido Teorema de Rao-Blackwell.

O Capítulo 3 é dedicado a técnicas de obtenção de estimadores, dentre as quais destacamos os métodos de máxima verossimilhança e dos momentos. Propriedades dos estimadores de máxima verossimilhança em grandes amostras são também consideradas. Essas propriedades permitem a realização de inferências em modelos mais complexos que são comumente utilizados em situações práticas.

No Capítulo 4 consideramos as idéias básicas da teoria das decisões, enfatizando a importância da função de risco como um meio de obtenção de bons estimadores. A utilização da função de risco permite a derivação de estimadores do tipo minimax e também de estimadores de Bayes, incorporando uma distribuição a priori para descrever conhecimentos subjetivos a cerca dos parâmetros de interesse.

A construção de intervalos de confiança com coeficientes de confiança exatos e aproximados é considerada no Capítulo 5. Um método importante de

construção de intervalos é o uso de quantidades pivotais. Tal enfoque propicia a construção de intervalos exatos para vários modelos importantes e aproximados em situações mais complexas. Intervalos Bayesianos baseados na distribuição a posteriori são também considerados.

O Capítulo 6 é dedicado à construção de testes de hipóteses. Testes ótimos para o caso de hipótese nula simples contra alternativa simples são derivados a partir do Lema de Neyman-Pearson. Algumas generalizações para hipóteses compostas são também consideradas. Problemas mais complexos que podem envolver hipóteses bilaterais são tratados utilizando a estatística da razão de verossimilhanças generalizada que, apesar de não possuir propriedades ótimas, leva em geral a bons procedimentos que não apresentam muita dificuldade de implementação.

Não incluímos no texto tabelas estatísticas, pois a ênfase maior é dada a problemas teóricos. No caso de haver necessidade de utilização de tabelas, sugerimos aos estudantes utilizar as tabelas em Bussab e Morettin (1987).

Agradecemos às colegas Elisete da Conceição Quintaneiro Aubin, Márcia D'Elia Branco e Sílvia Lopes de Paula Ferrari que leram as versões preliminares e contribuíram com várias sugestões. Agradecemos também à aluna Jacqueline Sant'Eufemia David pela elaboração das figuras.

São Paulo, setembro de 2000

Helena Bolfarine e Mônica C. Sandoval

# 1. Elementos Básicos

## 1.1 Alguns Modelos Especiais

Nesta seção consideramos alguns modelos probabilísticos que são comumente utilizados na análise de dados em problemas práticos. O modelo probabilístico (ou estatístico) é de suma importância para inferir resultados da amostra para a população toda. É importante que, na seleção do modelo a ser utilizado, o estatístico tenha em mente que o modelo deve representar, na medida do possível, a complexidade que envolve o mundo real da população em estudo. Entre os modelos mais utilizados, temos

### 1.1.1 O modelo normal

Dizemos que  $X$  tem distribuição normal com média  $\mu$  e variância  $\sigma^2$ , que denotamos por  $X \sim N(\mu, \sigma^2)$ , se a função de densidade de probabilidade 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  $\sigma^2 > 0$ . Nesse caso,  $\mu$  e  $\sigma^2$  são denominados parâmetros da distribuição e o suporte de  $X$ , isto é,  $A(x) = \{x, f(x) > 0\}$ , é a reta toda. Notemos também que

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

Situações práticas em que o modelo normal é comumente utilizado incluem características populacionais, tais como: peso, altura, pressão arterial, quociente de inteligência, etc.

### 1.1.2 O modelo exponencial

Dizemos que  $X$  tem distribuição exponencial com parâmetro  $\theta$ , que denotamos por  $X \sim Exp(\theta)$ , quando a função de densidade de probabilidade de  $X$  é dada por

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

em que  $\theta > 0$ . Nesse caso,  $A(x) = \{x, x > 0\}$ . Notemos também que

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

O modelo exponencial é comumente empregado para descrever tempo de vida de equipamentos. Lembremos que o modelo exponencial tem a bem conhecida propriedade da falta de memória, ou seja, se o tempo de vida de um equipamento segue a distribuição exponencial, então, em qualquer instante, o equipamento é como se fosse novo, não importando o quanto ele já tenha sido utilizado.

### 1.1.3 O modelo 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 função de probabilidade 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 discreto e é dado por  $A(x) = \{x, x = 0, 1, \dots, n\}$ . Temos também que

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

Lembremos que, se  $X$  tem distribuição  $\text{Binomial}(n, \theta)$ , então, podemos escrever  $X = Y_1 + \dots + Y_n$ , sendo  $Y_1, \dots, Y_n$   $n$  variáveis aleatórias independentes e de Bernoulli, ou seja, a função de probabilidade de  $Y_i$  é dada por

$$f(y_i|\theta) = \theta^{y_i} (1 - \theta)^{1-y_i}, \quad y_i = 0, 1,$$

$i = 1, \dots, n$ . O modelo binomial (ou de Bernoulli) é comumente empregado em situações em que associamos a cada observação da amostra dois tipos de resposta (como, por exemplo, sim e não, ou sucesso e fracasso) aos quais associamos os valores 0 e 1. Tais situações envolvem, por exemplo, pesquisas eleitorais, em que os indivíduos na população são ou não favoráveis a determinado partido ou candidato; proporção de peças defeituosas produzidas em uma linha de produção e assim por diante.

### 1.1.4 O modelo de Poisson

Um outro modelo comumente empregado na prática é o modelo de Poisson. Dizemos que a variável aleatória  $X$  tem distribuição de Poisson com parâmetro



$\theta$ , que denotamos por  $X \sim \text{Poisson}(\theta)$ , quando a função de probabilidade é dada por

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

em que  $\theta > 0$ . Nesse caso, o suporte de  $X$  é o conjunto  $A(x) = \{x, x = 0, 1, \dots\}$ . Temos também que,

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

O modelo de Poisson é bastante utilizado para descrever situações que envolvem, 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.

### 1.1.5 O modelo uniforme

O modelo uniforme é bastante importante do ponto de vista teórico. Dizemos que  $X$  tem distribuição uniforme no intervalo  $(0, \theta)$ , que denotamos por  $X \sim U(0, \theta)$ , se a função de densidade de  $X$  é dada por

$$\begin{aligned} f(x|\theta) &= \begin{cases} \frac{1}{\theta}, & 0 < x < \theta, \\ 0, & \text{caso contrário,} \end{cases} \\ &= \frac{1}{\theta} I_{(0, \theta)}(x), \end{aligned}$$

$\theta > 0$ , em que

$$I_{(0, \theta)}(x) = \begin{cases} 1, & 0 < x < \theta, \\ 0, & \text{caso contrário,} \end{cases}$$

ou seja,  $I_{(0, \theta)}(x)$  é a função indicadora do intervalo  $(0, \theta)$ . Notemos que, nesse caso,  $A(x) = \{x, 0 < x < \theta\}$ , ou seja, o suporte da variável  $X$  (ou de  $f(x|\theta)$ ) depende do parâmetro  $\theta$ . No caso dos modelos normal, exponencial, binomial e de Poisson, isso não acontece, ou seja, nesses casos, o suporte da distribuição de  $X$  é independente de  $\theta$ . Temos também que, se  $X \sim U(0, \theta)$ , então,

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

No decorrer do texto, outros modelos paramétricos, como por exemplo, o modelo uniforme discreto e o modelo gama, serão apresentados. Veremos também que os modelos normal, exponencial, binomial e de Poisson são membros de uma família bastante geral de modelos, que é a família exponencial.

## 1.2 Tipos de Problemas

No presente texto, vamos nos ater exclusivamente a problemas de estimação e de testes de hipóteses.

**Definição 1.2.1.** *Seja  $X$  uma variável aleatória com função de densidade (ou de probabilidade) que abreviamos por f.d.p. (f.p.) e que denotamos por  $f(x|\theta)$ , em que  $\theta$  é um parâmetro desconhecido. 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 distribuição da variável aleatória  $X$  pertence a certa família de distribuições em que um particular elemento é especificado, quando o valor do parâmetro  $\theta$  é especificado.

No caso de um problema de **estimação**, o objetivo é procurar, segundo algum critério especificado, valores que representem adequadamente os parâmetros desconhecidos. No caso de problemas de **testes de hipóteses**, o objetivo é verificar a validade de afirmações sobre um valor (ou valores) do(s) parâmetro(s) desconhecido(s). Por exemplo, quando o interesse é verificar se a proporção  $\theta$  de eleitores de determinado candidato é maior que  $1/2$  (ou 50%), as hipóteses a serem testadas são  $H_0 : \theta \leq 1/2$  versus  $H_1 : \theta > 1/2$ . Quando estamos interessados em verificar se o peso médio,  $\mu$ , de pacotes de um quilograma empacotados por determinada máquina realmente é um quilograma, então, as hipóteses a serem testadas podem ser representadas por  $H_0 : \mu = 1$  versus  $H_1 : \mu \neq 1$ .

## 1.3 Amostras, Estatísticas e Estimadores

Nesta seção os conceitos de estatística e estimador são introduzidos. Critérios para a comparação de estimadores são também considerados.

**Definição 1.3.1.** *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. De maneira mais formal, temos

**Definição 1.3.2.** *Uma sequência  $X_1, \dots, X_n$  de  $n$  variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com função de densidade (f.d.p.) ou, no caso discreto, função de probabilidade (f.p.)  $f(x|\theta)$  é dita ser uma amostra aleatória de tamanho  $n$  da distribuição de  $X$ . Nesse caso, temos,*

$$(1.3.1) \quad f(x_1, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta) = f(x_1 | \theta) \dots f(x_n | \theta).$$

Concluimos, a partir da Definição 1.3.2, que usamos a amostra  $X_1, \dots, X_n$  para obter informação sobre o parâmetro  $\theta$ . A função de densidade (ou de probabilidade) conjunta dada em (1.3.1) é denominada **função de verossimilhança** de  $\theta$ , correspondente à amostra observada  $\mathbf{x} = (x_1, \dots, x_n)'$  e será denotada por

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

**Definição 1.3.3.** *Qualquer função da amostra que não depende de parâmetros desconhecidos é denominada uma estatística.*

No exemplo que apresentamos a seguir, consideramos várias estatísticas que serão utilizadas com frequência nos capítulos seguintes.

**Exemplo 1.3.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$ , com f.d.p. ou f.p.  $f(x|\theta)$ . Exemplos de estatísticas são

- (i)  $X_{(1)} = \min(X_1, \dots, X_n)$ ,
- (ii)  $X_{(n)} = \max(X_1, \dots, X_n)$ ,
- (iii)  $\tilde{X} = \text{med}(X_1, \dots, X_n)$ ,
- (iv)  $\overline{X} = \frac{1}{n} \sum_{i=1}^n X_i$ ,
- (v)  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2$ .

Em (i), (ii) e (iii) acima,  $\min(\cdot)$ ,  $\max(\cdot)$  e  $\text{med}(\cdot)$  denotam, respectivamente, o mínimo, o máximo e a mediana amostral observada. Por outro lado,  $\overline{X}$  e  $\hat{\sigma}^2$  denotam, respectivamente, a média e a variância amostrais.

**Definição 1.3.4.** *O conjunto  $\Theta$  em que  $\theta$  toma valores é denominado espaço paramétrico.*

**Exemplo 1.3.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\mu, \sigma^2)$ .

- (i) Se  $\sigma^2 = 1$ , então  $\theta = \mu$  é o parâmetro desconhecido e

$$\Theta = \{\mu, -\infty < \mu < \infty\};$$

- (ii) Se  $\mu = 0$ , então  $\theta = \sigma^2$  é o parâmetro desconhecido e

$$\Theta = \{\sigma^2, \sigma^2 > 0\};$$

(iii) Se  $\mu$  e  $\sigma^2$  são desconhecidos então  $\theta = (\mu, \sigma^2)$  e

$$\Theta = \{(\mu, \sigma^2), -\infty < \mu < \infty \text{ e } \sigma^2 > 0\}.$$

**Definição 1.3.5.** *Qualquer estatística que assuma valores em  $\Theta$  é um estimador para  $\theta$ .*

Em muitas situações, o interesse é estimar uma função  $g(\theta)$ . Suponha, por exemplo, que no caso (iii) do exemplo anterior, o objetivo é estimar somente  $\mu$ , sendo  $\sigma^2$  um parâmetro de perturbação. Nesse caso,  $g(\theta) = \mu$ .

**Definição 1.3.6.** *Qualquer estatística que assuma valores somente no conjunto dos possíveis valores de  $g(\theta)$  é um estimador para  $g(\theta)$ .*

Um dos grandes problemas da estatística é o de encontrar um estimador razoável para o parâmetro desconhecido  $\theta$  ou para uma função  $g(\theta)$ . Um dos procedimentos comumente utilizados para se avaliar o desempenho de um estimador é o seu erro quadrático médio que é considerado a seguir.

**Definição 1.3.7.** *O erro quadrático médio (EQM) de um estimador  $\hat{\theta}$  do parâmetro  $\theta$  é dado por*

$$EQM[\hat{\theta}] = E[(\hat{\theta} - \theta)^2].$$

Pode-se mostrar (ver Exercício 1.1) que

$$(1.3.2) \quad EQM[\hat{\theta}] = Var[\hat{\theta}] + B^2(\hat{\theta}),$$

em que

$$B(\hat{\theta}) = E[\hat{\theta}] - \theta$$

é denominado o vício do estimador  $\hat{\theta}$ . Dizemos que um estimador  $\hat{\theta}$  é **não viciado** para  $\theta$  se

$$E[\hat{\theta}] = \theta,$$

para todo  $\theta \in \Theta$ , ou seja  $B(\hat{\theta}) = 0$ , para todo  $\theta \in \Theta$ . Se  $\lim_{n \rightarrow \infty} B(\hat{\theta}) = 0$  para todo  $\theta \in \Theta$ , dizemos que o estimador  $\hat{\theta}$  é **assintoticamente não viciado** para  $\theta$ . No caso em que  $\hat{\theta}$  é um estimador não viciado para  $\theta$ , temos que

$$EQM[\hat{\theta}] = Var[\hat{\theta}],$$

ou seja, o erro quadrático médio de  $\hat{\theta}$  se reduz à sua variância. Um outro conceito importante em grandes amostras ( $n \rightarrow \infty$ ) é a propriedade de consistência que será considerada na Seção 3.7.

**Exemplo 1.3.3.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com  $E[X] = \mu$  e  $Var[X] = \sigma^2$ . Temos, então, 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

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

Portanto  $\bar{X}$  é um estimador não viciado para  $\mu$ . Com relação à variância amostral, temos

$$\begin{aligned} E[\hat{\sigma}^2] &= \frac{1}{n} E \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n E[(X_i - \bar{X})^2] \\ &= \frac{1}{n} \sum_{i=1}^n E\{[(X_i - \mu) - (\bar{X} - \mu)]^2\} \end{aligned}$$

$$(1.3.3) \quad = \frac{(n-1)}{n} \sigma^2.$$

Portanto  $\hat{\sigma}^2$  é viciado para  $\sigma^2$ , mas é assintoticamente não viciado, ou seja, à medida que o tamanho da amostra aumenta, o vício diminui.

O erro quadrático médio é comumente empregado na comparação de estimadores. Dizemos, então, que  $\hat{\theta}_1$  é **melhor** que  $\hat{\theta}_2$  se

$$(1.3.4) \quad EQM[\hat{\theta}_1] \leq EQM[\hat{\theta}_2],$$

para todo  $\theta$ , com  $\leq$  substituído por  $<$  pelo menos para um valor de  $\theta$ . Nesse caso, o estimador  $\hat{\theta}_2$  é dito ser **inadmissível**. Se existir um estimador  $\hat{\theta}^*$  tal que para todo estimador  $\hat{\theta}$  de  $\theta$  com  $\hat{\theta} \neq \hat{\theta}^*$

$$(1.3.5) \quad EQM[\hat{\theta}^*] \leq EQM[\hat{\theta}],$$

para todo  $\theta$  com  $\leq$  substituído por  $<$  para pelo menos um  $\theta$ , então  $\hat{\theta}^*$  é dito ser ótimo para  $\theta$ . Notemos que, se em (1.3.5) os estimadores são não viciados, então  $\hat{\theta}^*$  é dito ser o estimador não viciado de variância uniformemente mínima, se

$$Var[\hat{\theta}^*] \leq Var[\hat{\theta}],$$

para todo  $\theta$ , com  $\leq$  substituído por  $<$  para pelo menos um  $\theta$ .

**Exemplo 1.3.4.** Sejam  $X_1, X_2, X_3$  uma amostra aleatória da variável aleatória  $X$  com  $E[X] = \theta$  e  $Var[X] = 1$ . Consideremos os estimadores

$$\hat{\theta}_1 = \bar{X} = \frac{X_1 + X_2 + X_3}{3} \quad \text{e} \quad \hat{\theta}_2 = \frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3.$$

Como no Exemplo 1.3.3,

$$E[\hat{\theta}_1] = \theta \quad \text{e} \quad \text{Var}[\hat{\theta}_1] = \frac{1}{3}.$$

Temos também (ver Exercício 1.3) que

$$(1.3.6) \quad E[\hat{\theta}_2] = \theta \quad \text{e} \quad \text{Var}[\hat{\theta}_2] = \frac{6}{16}.$$

Como  $\hat{\theta}_1$  e  $\hat{\theta}_2$  são ambos não viciados, segue de (1.3.4) que  $\bar{X}$  é melhor que  $\hat{\theta}_2$ , pois  $\text{Var}[\bar{X}] < \text{Var}[\hat{\theta}_2]$ , para todo  $\theta$ .

**Exemplo 1.3.5.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com  $E[X] = \theta$  e  $\text{Var}[X] = \sigma^2$ , em que  $\sigma^2$  é conhecido. Consideramos agora os estimadores lineares

$$X_L = \sum_{i=1}^n l_i X_i,$$

em que  $l_i \geq 0$ ,  $i = 1, \dots, n$  são constantes conhecidas. Como

$$E[X_L] = E\left[\sum_{i=1}^n l_i X_i\right] = \sum_{i=1}^n l_i E[X_i] = \theta \sum_{i=1}^n l_i,$$

temos que  $X_L$  é um estimador não viciado para  $\theta$  se e somente se

$$(1.3.7) \quad \sum_{i=1}^n l_i = 1.$$

O estimador  $X_L$  com a condição (1.3.7) é então uma combinação linear convexa de  $X_1, \dots, X_n$ . Notemos que  $\hat{\theta}_1$  e  $\hat{\theta}_2$  considerados no Exemplo 1.3.4 são combinações lineares convexas de  $X_1, X_2, X_3$ . Temos também que

$$(1.3.8) \quad \text{Var}[X_L] = \sum_{i=1}^n l_i^2 \text{Var}[X_i] = \sigma^2 \sum_{i=1}^n l_i^2.$$

Portanto o estimador  $X_L$ , que é não viciado e apresenta a menor variância, é obtido minimizando-se (1.3.8) sujeito à condição (1.3.7). Para atingir tal objetivo, sendo  $\bar{l} = \sum_{i=1}^n l_i/n = 1/n$  a média dos  $l_i$ 's, temos que

$$\sum_{i=1}^n (l_i - \bar{l})^2 = \sum_{i=1}^n l_i^2 - n\bar{l}^2 = \sum_{i=1}^n l_i^2 - 1/n,$$

de modo que

$$\begin{aligned}
 \text{Var}[X_L] &= \sigma^2 \sum_{i=1}^n l_i^2 \\
 (1.3.9) \quad &= \sigma^2 \left\{ \sum_{i=1}^n \left( l_i - \frac{1}{n} \right)^2 + \frac{1}{n} \right\}.
 \end{aligned}$$

Assim, a expressão (1.3.9) será mínima quando  $l_i = 1/n$ , ou seja o estimador  $X_L$  com menor variância é a média amostral  $\bar{X}$ . Portanto, dentre todos os estimadores lineares não viciados  $X_L$ , o que apresenta a menor variância é a média amostral  $\bar{X}$ . De (1.3.9) segue também que  $\text{Var}[\bar{X}] = \sigma^2/n$ . Uma outra forma de minimizar a variância (1.3.8), sob a condição (1.3.7), é feita utilizando-se de multiplicadores de Lagrange. Nesse caso, temos o "Lagrangeano"

$$\mathcal{L}(\lambda) = \sigma^2 \sum_{i=1}^n l_i^2 - \lambda \left( \sum_{i=1}^n l_i - 1 \right).$$

Derivando sucessivamente com relação a  $l_1, \dots, l_n$ , temos as equações

$$2\sigma^2 l_1 - \lambda = 0, \quad \dots, \quad 2\sigma^2 l_n - \lambda = 0,$$

de modo que

$$2l_i \sigma^2 = 2l_n \sigma^2,$$

logo

$$l_i = l_n,$$

$i = 1, \dots, n$ . Sendo  $\sum_{i=1}^n l_i = 1$ , segue que  $l_i = 1/n$ ,  $i = 1, \dots, n$ , como concluído acima.

**Exemplo 1.3.6.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\mu, \sigma^2)$ . Conforme visto no Exemplo 1.3.3,  $\hat{\sigma}^2$  é um estimador viciado para  $\sigma^2$ . De (1.3.3) segue que

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

é um estimador não viciado para  $\sigma^2$ . Por outro lado, temos (ver Exercício 1.4) que

$$(1.3.10) \quad EQM[S^2] = \text{Var}[S^2] = \frac{2\sigma^4}{n-1},$$

e que

$$(1.3.11) \quad EQM[\hat{\sigma}^2] = \frac{2\sigma^4}{(n-1)} \left[ 1 - \frac{(3n-1)}{2n^2} \right].$$

Notemos que  $\hat{\sigma}^2$ , apesar de viciado, apresenta um  $EQM$  menor que o  $EQM$  do estimador  $S^2$ .

**Exemplo 1.3.7.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$ , com distribuição de Bernoulli com parâmetro  $\theta$ , ou seja  $Binomial(1, \theta)$ . Conforme visto no modelo binomial,  $Y = X_1 + \dots + X_n$  tem distribuição  $Binomial(n, \theta)$ . Consideremos os estimadores

$$\hat{\theta}_1 = \bar{X} = \frac{Y}{n} \quad \text{e} \quad \hat{\theta}_2 = \frac{Y + \sqrt{n}/2}{n + \sqrt{n}}.$$

Como  $E[\bar{X}] = \theta$ , temos que

$$EQM[\hat{\theta}_1] = Var[\bar{X}] = \frac{\theta(1-\theta)}{n}.$$

Por outro lado,

$$E[\hat{\theta}_2] = E\left[\frac{Y + \sqrt{n}/2}{n + \sqrt{n}}\right] = \frac{n\theta + \sqrt{n}/2}{n + \sqrt{n}} = \frac{n}{n + \sqrt{n}}\theta + \frac{\sqrt{n}/2}{n + \sqrt{n}},$$

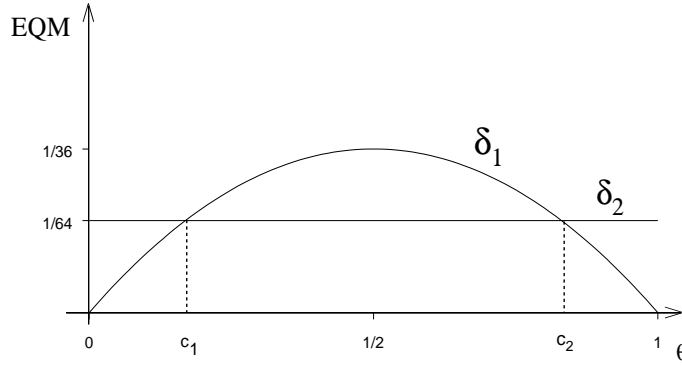
de modo que  $\hat{\theta}_2$  é um estimador viciado para  $\theta$ . Notemos que, na verdade, o vício é uma função linear de  $\theta$ . Portanto

$$\begin{aligned} EQM[\hat{\theta}_2] &= E\left[\left(\frac{Y + \sqrt{n}/2}{n + \sqrt{n}} - \theta\right)^2\right] \\ &= \frac{1}{(n + \sqrt{n})^2} E\left\{\left[(Y - n\theta) + \sqrt{n}\left(\frac{1}{2} - \theta\right)\right]^2\right\} \\ &= \frac{1}{(n + \sqrt{n})^2} \left\{Var[Y] + n\left(\frac{1}{2} - \theta\right)^2\right\} \\ &= \frac{n}{4(n + \sqrt{n})^2}. \end{aligned}$$

Um fato importante a ser notado é que o  $EQM$  do estimador  $\hat{\theta}_2$  é independente de  $\theta$ . O  $EQM$  dos dois estimadores é representado graficamente na Figura 1.1, para  $n = 9$ .

Temos, então, que nenhum dos estimadores é melhor uniformemente, isto é, para todo  $\theta$ . Para  $c_1 < \theta < c_2$ ,  $EQM[\hat{\theta}_2] < EQM[\hat{\theta}_1]$ , ou seja,  $\hat{\theta}_2$  é melhor que  $\hat{\theta}_1$ . Por outro lado, para  $\theta < c_1$  ou  $\theta > c_2$ , temos que  $EQM[\hat{\theta}_1] < EQM[\hat{\theta}_2]$ , ou seja,  $\hat{\theta}_1$  é melhor que  $\hat{\theta}_2$ . Para o cálculo de  $c_1$  e  $c_2$ , ver Exercício 1.5.



**Figura 1.1.** *EQM* de  $\delta_1 = \hat{\theta}_1$  e  $\delta_2 = \hat{\theta}_2$ 

**Exemplo 1.3.8.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim U(0, \theta)$ . Vamos considerar  $\hat{\theta}_1 = \bar{X}$  e  $\hat{\theta}_2 = X_{(n)}$  como estimadores de  $\theta$ . Como  $E[X] = \theta/2$  e  $Var[X] = \theta^2/12$  (ver o modelo (1.1.4)), temos que

$$(1.3.12) \quad E[\hat{\theta}_1] = E[\bar{X}] = \frac{\theta}{2},$$

e

$$(1.3.13) \quad Var[\hat{\theta}_1] = \frac{\theta^2}{12n}.$$

Portanto o estimador  $\hat{\theta}_1$  é viciado para  $\theta$ . Combinando (1.3.12) e (1.3.13) em (1.3.2), temos que

$$EQM[\hat{\theta}_1] = \frac{\theta^2}{12n} + \left(\frac{\theta}{2} - \theta\right)^2 = \frac{(1+3n)}{12n}\theta^2.$$

Por outro lado, a função de densidade de  $X_{(n)}$  (ver Exercício 1.6) é dada por

$$(1.3.14) \quad f_{X_{(n)}}(x|\theta) = \frac{nx^{n-1}}{\theta^n}, \quad 0 < x < \theta,$$

logo

$$(1.3.15) \quad E[X_{(n)}] = \frac{n}{n+1}\theta \quad \text{e} \quad Var[X_{(n)}] = \frac{n\theta^2}{(n+1)^2(n+2)}.$$

Portanto

$$EQM[\hat{\theta}_2] = \frac{n\theta^2}{(n+1)^2(n+2)} + \frac{\theta^2}{(n+1)^2} = \frac{2\theta^2}{(n+1)(n+2)}.$$

A Tabela 1.1 mostra o valor do  $EQM$  dos dois estimadores para vários valores de  $n$ . Notemos também que, quando  $n \rightarrow \infty$ ,  $EQM[\hat{\theta}_1] \rightarrow \theta^2/4$  e que  $EQM[\hat{\theta}_2] \rightarrow 0$ .

**Tabela 1.1.**  $EQM$  de  $\hat{\theta}_1$  e  $\hat{\theta}_2$

$n$	$EQM[\hat{\theta}_1]$	$EQM[\hat{\theta}_2]$	$EQM[\hat{\theta}_2]/EQM[\hat{\theta}_1]$
3	$5\theta^2/18$	$\theta^2/10$	0,27
5	$4\theta^2/15$	$\theta^2/21$	0,12
10	$31\theta^2/120$	$\theta^2/662$	0,04
20	$61\theta^2/240$	$\theta^2/2312$	0,01

Portanto  $X_{(n)}$  é melhor que  $\bar{X}$  para todo  $\theta$  e  $n > 1$ .

**Exemplo 1.3.9.** Consideremos uma urna com  $N$  bolas idênticas marcadas com os números  $1, \dots, N$ . O objetivo é a estimação de  $N$ , o número de bolas numeradas na urna. Esse problema está muitas vezes associado ao problema da estimação do número  $N$  de táxis em uma cidade, em que os táxis estão numerados de 1 a  $N$ . Portanto uma determinada quantidade ( $n$ ) de bolas (táxis) é observada, com reposição. Associada à  $i$ -ésima observação, temos a variável aleatória

$X_i$ : número da  $i$ -ésima bola (táxi) retirada da urna,

$i = 1, \dots, n$ . Nesse caso,

$$P[X_i = k] = \frac{1}{N}, \quad k = 1, \dots, N.$$

Portanto a distribuição de  $X_i$  é uniforme discreta, pois a distribuição de  $X_i$  associa a mesma probabilidade a todos os possíveis valores de  $X_i$ ,  $i = 1, \dots, n$ . Como possíveis estimadores de  $N$ , consideremos inicialmente  $\hat{N}_1 = \bar{X}$  e  $\hat{N}_2 = X_{(n)}$ . Não é difícil verificar que

$$(1.3.16) \quad E[\hat{N}_1] = E[\bar{X}] = \frac{N+1}{2}.$$

Por outro lado, desde que

$$P[X_{(n)} = k] = P[X_{(n)} \leq k] - P[X_{(n)} \leq k-1] = \left(\frac{k}{N}\right)^n - \left(\frac{k-1}{N}\right)^n,$$

temos que

$$E[X_{(n)}] = N^{-n} \left[ N^{n+1} - \sum_{k=1}^N (k-1)^n \right].$$

Usando a aproximação (Feller, 1976),

$$\sum_{k=1}^N (k-1)^n = 1^n + \dots + (N-1)^n \cong \int_0^N y^n dy = \frac{N^{n+1}}{n+1},$$

(para  $N$  grande), temos que

$$(1.3.17) \quad E[\hat{N}_2] = E[X_{(n)}] \cong N^{-n} \left[ N^{n+1} - \frac{N^{n+1}}{n+1} \right] = \frac{n}{n+1} N.$$

De (1.3.16) e (1.3.17), podemos definir novos estimadores. Por exemplo,

$$\hat{N}_3 = 2\bar{X} - 1,$$

que é não viciado e

$$\hat{N}_4 = \frac{n+1}{n} X_{(n)},$$

que é aproximadamente não viciado. Se  $n = 8$  bolas são retiradas com reposição da caixa e os números observados são: 124, 212, 315, 628, 684, 712, 782, 926, então,  $\hat{N}_1 = \bar{X} = 547,875$ ,  $\hat{N}_3 = 2\bar{X} - 1 = 1095$ ,  $\hat{N}_2 = X_{(n)} = 926$ , e  $\hat{N}_4 = 1042$ . Podemos considerar também o estimador

$$\hat{N}_5 = \frac{X_{(n)}^{n+1} - (X_{(n)} - 1)^{n+1}}{X_{(n)}^n - (X_{(n)} - 1)^n},$$

que é um estimador não viciado para  $N$  (ver Exercício 1.7).

## 1.4 Exercícios

**1.1.** Verifique a validade da expressão (1.3.2).

**1.2.** Verifique a validade da expressão (1.3.3).

**1.3.** Verifique a validade da expressão (1.3.6).

**1.4.** Verifique a validade das expressões (1.3.10) e (1.3.11).

**1.5.** Encontre  $c_1$  e  $c_2$  na Figura 1.1. que são os pontos de intersecção dos erros quadráticos médios de  $\hat{\theta}_1$  e  $\hat{\theta}_2$ .

**1.6.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim U(0, \theta)$ . Mostre que a função de densidade de probabilidade de  $X_{(n)}$  é como dada em (1.3.14), com esperança e variância como dadas em (1.3.15).

**1.7.** Mostre que o  $\hat{N}_5$  no Exemplo 1.3.9 é um estimador não viciado para  $N$ .

**1.8.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição da variável aleatória  $X$ , em que  $X \sim N(\mu, 1)$ . Considere os estimadores  $\hat{\mu}_1 = \bar{X}$  e  $\hat{\mu}_2 = 10$ . Encontre o *EQM* de  $\hat{\mu}_1$  e de  $\hat{\mu}_2$  como função de  $\mu$ . Faça um gráfico do *EQM* para  $n = 10$ .

**1.9.** Seja  $X$  uma única variável aleatória com distribuição de Bernoulli com parâmetro  $\theta$ . Sejam  $\hat{\theta}_1 = X$  e  $\hat{\theta}_2 = 1/2$  dois estimadores de  $\theta$ .

- (i) Verifique se  $\hat{\theta}_1$  e  $\hat{\theta}_2$  são não viciados para  $\theta$ .
- (ii) Compare os *EQMs*. Faça um gráfico dos *EQMs* como função de  $\theta$ .

**1.10.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição da variável aleatória  $X$  com f.d.p. dada por

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

- (i) Especifique o espaço paramétrico e o suporte associado à distribuição de  $X$ .
- (ii) Verifique se  $\hat{\theta}_1 = \bar{X}$  e  $\hat{\theta}_2 = X_{(1)}$  são estimadores não viciados para  $\theta$ .
- (iii) Encontre e compare os *EQMs* dos dois estimadores. Faça um gráfico como função de  $\theta$ .

**1.11.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição da variável aleatória  $X$  com f.d.p. dada por

$$f(x|\theta) = \frac{2x}{\theta^2}, \quad 0 < x < \theta, \quad \theta > 0.$$

- (i) Especifique o espaço paramétrico e o suporte associado à distribuição de  $X$ .
- (ii) Verifique se  $\hat{\theta}_1 = \bar{X}$  e  $\hat{\theta}_2 = X_{(n)}$  são não viciados para  $\theta$ .
- (iii) Encontre e compare os *EQMs* dos dois estimadores. Faça um gráfico dos *EQMs* como função de  $\theta$ .

**1.12.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição de uma variável aleatória  $X \sim U(0, \theta)$ . Considere os estimadores  $\hat{\theta}_1 = c_1 \bar{X}$  e  $\hat{\theta}_2 = c_2 X_{(n)}$ .

- (i) Encontre  $c_1$  e  $c_2$  que tornam os estimadores não viciados.
- (ii) Encontre e compare os *EQMs* dos dois estimadores.

**1.13.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição da variável aleatória  $X \sim N(0, \sigma^2)$ . Seja  $S^2 = \sum_{i=1}^n X_i^2$ . Considere os estimadores

$$\hat{\sigma}_c^2 = cS^2.$$

- (i) Encontre o *EQM* do estimador acima.
- (ii) Encontre o valor de  $c$  que minimiza o *EQM* em (i).

## 2. Estimadores Eficientes e Estatísticas Suficientes

Neste capítulo será apresentada a noção de estimador eficiente, como sendo aquele que atinge o limite inferior da variância dos estimadores não viciados. Estimadores eficientes são obtidos apenas para distribuições que são membros de uma classe especial, que é a família exponencial de distribuições. Veremos também que todo estimador para ser ótimo, segundo o critério do menor erro quadrático médio, deve ser função de uma estatística suficiente. De modo informal, estatísticas suficientes para um parâmetro (ou para uma distribuição) são aquelas que condensam os dados sem perder nenhuma informação contida nos mesmos. Ou seja, elas são tão informativas para o parâmetro (ou para a distribuição) quanto a amostra toda.

### 2.1 Estimadores Eficientes

Eficiência de um estimador  $\hat{\theta}$  de um parâmetro  $\theta$  é definida a seguir.

**Definição 2.1.1.** Chamamos de eficiência de um estimador  $\hat{\theta}$ , não viciado para o parâmetro  $\theta$ , o quociente

$$e(\hat{\theta}) = \frac{LI(\theta)}{Var[\hat{\theta}]},$$

onde  $LI(\theta)$  é o limite inferior da variância dos estimadores não viciados de  $\theta$ .

Convém notar que:

(i)  $e(\hat{\theta}) = 1$  quando  $LI(\theta) = Var[\hat{\theta}]$ , ou seja, quando a variância de  $\hat{\theta}$  coincide com o limite inferior da variância dos estimadores não viciados de  $\theta$ . Nesse caso,  $\hat{\theta}$  é dito ser **eficiente**;

(ii) como veremos no teorema seguinte,

$$(2.1.1) \quad LI(\theta) = \frac{1}{nE \left[ \left( \frac{\partial \log f(X|\theta)}{\partial \theta} \right)^2 \right]},$$

quando certas condições de regularidade estão satisfeitas;

(iii) as condições de regularidade a que nos referimos no item (ii) são basicamente duas, isto é, que o suporte  $A(x) = \{x, f(x|\theta) > 0\}$  seja independente de  $\theta$  e que seja possível a troca das ordens das operações de derivação e de integração sob a distribuição da variável aleatória  $X$ ;

(iv) a não ser que mencionado o contrário, todo logaritmo utilizado no texto é calculado na base  $e$ .

**Exemplo 2.1.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\mu, \sigma^2)$ , em que  $\sigma^2$  é conhecido. Temos que

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

e

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

Portanto

$$(2.1.2) \quad \frac{\partial \log f(x|\mu)}{\partial \mu} = \frac{(x-\mu)}{\sigma^2}.$$

Assim,

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

logo concluímos, juntamente com (2.1.1), que

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

Conforme visto no Exemplo 1.3.3, temos que

$$Var[\bar{X}] = \frac{\sigma^2}{n} = LI(\mu),$$

de modo que  $\bar{X}$  é um estimador eficiente para  $\mu$ . De (2.1.2), temos também que

$$(2.1.3) \quad E \left[ \frac{\partial \log f(X|\mu)}{\partial \mu} \right] = \frac{1}{\sigma^2} E[X - \mu] = 0.$$

**Definição 2.1.2.** A quantidade

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

é chamada função escore.

O resultado (2.1.3) na verdade vale em geral quando valem as condições de regularidade, ou seja,

$$(2.1.4) \quad E \left[ \frac{\partial \log f(X|\theta)}{\partial \theta} \right] = 0.$$

Portanto o valor esperado da função escore é sempre igual a zero.

**Definição 2.1.3.** *A quantidade*

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

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

Como consequência de (2.1.4) temos que

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

pois para uma variável aleatória  $X$  qualquer com  $E[X] = 0$ ,  $Var[X] = E[X^2]$ . Um resultado importante (veja o Exercício 2.6) 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 importante 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, ou seja, sendo

$$(2.1.5) \quad L(\theta; \mathbf{x}) = f(x_1, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta),$$

a densidade conjunta de  $X_1, \dots, X_n$ , temos que

$$(2.1.6) \quad \begin{aligned} & E \left[ \left( \frac{\partial \log L(\theta; \mathbf{X})}{\partial \theta} \right)^2 \right] = -E \left[ \frac{\partial^2 \log L(\theta; \mathbf{X})}{\partial \theta^2} \right] \\ & = -E \left[ \sum_{i=1}^n \frac{\partial^2 \log f(X_i|\theta)}{\partial \theta^2} \right] = \sum_{i=1}^n E \left[ -\frac{\partial^2 \log f(X_i|\theta)}{\partial \theta^2} \right] = nI_F(\theta), \end{aligned}$$

pois  $X_i$ ,  $i = 1, \dots, n$  têm a mesma informação que  $X$ . Lembremos que, sendo  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$ , então  $X_1, \dots, X_n$  são independentes e identicamente distribuídas com a mesma distribuição que  $X$ .

**Teorema 2.1.1. 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*

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

**Prova.** Vamos considerar o caso em que  $X$  é uma variável aleatória contínua. Sendo  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$ , temos que

$$(2.1.7) \quad \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} L(\theta; \mathbf{x}) dx_1 \dots dx_n = 1,$$

onde  $L(\theta; \mathbf{x})$  é dada em (2.1.5). Desde que  $\hat{\theta}$  é não viciado, ou seja,  $E[\hat{\theta}] = \theta$ , temos também que

$$(2.1.8) \quad \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \hat{\theta} L(\theta; \mathbf{x}) dx_1 \dots dx_n = \theta.$$

Derivando ambos os lados de (2.1.7) com relação a  $\theta$ , temos que

$$\frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} L(\theta; \mathbf{x}) dx_1 \dots dx_n = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{\partial L(\theta; \mathbf{x})}{\partial \theta} dx_1 \dots dx_n = 0.$$

Por outro lado, de (2.1.8), temos que

$$\frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \hat{\theta} L(\theta; \mathbf{x}) dx_1 \dots dx_n = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \hat{\theta} \frac{\partial L(\theta; \mathbf{x})}{\partial \theta} dx_1 \dots dx_n = 1.$$

Como

$$\frac{\partial L(\theta; \mathbf{x})}{\partial \theta} = t(\theta; \mathbf{x}) L(\theta; \mathbf{x}),$$

onde

$$t(\theta; \mathbf{x}) = \frac{\partial \log L(\theta; \mathbf{x})}{\partial \theta},$$

temos das expressões acima que

$$E[t(\theta; \mathbf{X})] = 0,$$

e



$$E[\hat{\theta}t(\theta; \mathbf{X})] = 1.$$

Como

$$\rho_{\hat{\theta}t} = \frac{E[\hat{\theta}t(\theta; \mathbf{X})] - E[\hat{\theta}]E[t(\theta; \mathbf{X})]}{\sqrt{\text{Var}[\hat{\theta}]\text{Var}[t(\theta; \mathbf{X})]}},$$

onde  $\rho_{\hat{\theta}t}$  denota o coeficiente de correlação entre  $\hat{\theta}$  e  $t$ , de tal forma que  $\rho_{\hat{\theta}t}^2 \leq 1$ , temos que

$$\text{Var}[\hat{\theta}] \geq \frac{1}{\text{Var}[t(\theta; \mathbf{X})]}.$$

Como as variáveis  $X_1, \dots, X_n$  são independentes e identicamente distribuídas com densidade  $f(x|\theta)$ , temos de (2.1.5) e de (2.1.6) que

$$\text{Var}[t(\theta; \mathbf{X})] = \text{Var}\left[\frac{\partial \log L(\theta; \mathbf{X})}{\partial \theta}\right] = nI_F(\theta),$$

o que prova o resultado.

**Exemplo 2.1.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X \sim \text{Poisson}(\theta)$ , com função de probabilidade dada por

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

Nesse caso, temos que

$$\log f(x|\theta) = -\log x! + x \log \theta - \theta,$$

de modo que

$$\frac{\partial \log f(x|\theta)}{\partial \theta} = \frac{x}{\theta} - 1,$$

ou seja,

$$E\left[\frac{\partial^2 \log f(X|\theta)}{\partial \theta^2}\right] = -\frac{1}{\theta}.$$

Portanto

$$LI(\theta) = \frac{\theta}{n}.$$

Como  $\text{Var}[\bar{X}] = \theta/n$ , concluímos que  $\bar{X}$  é um estimador eficiente para  $\theta$ .

Enfatizamos que a desigualdade da informação (inicialmente chamada de Cramér-Rao) não é um método de construção de estimadores. Ela apenas possibilita verificar se determinado estimador é ou não eficiente. É então importante que sejam estabelecidos métodos para construção de estimadores que tenham alguma propriedade interessante, ou que levem a estimadores com “boas” propriedades. Contudo, antes de estabelecermos tais métodos (ou critérios), vamos considerar estatísticas que reduzam (condensem) os dados sem que haja perda de informação. Tais estatísticas são conhecidas como estatísticas suficientes.

## 2.2 Estatísticas Suficientes

Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com função de densidade ou de probabilidade  $f(x|\theta)$ . Quando resumimos a informação que os dados contêm sobre  $\theta$ , utilizando uma estatística, é importante que não haja perda de informação sobre  $\theta$ . Ou seja, a estatística a ser considerada deve, dentro do possível, conter toda a informação sobre  $\theta$  presente na amostra. Em outras palavras, se pudermos usar uma estatística  $T = T(X_1, \dots, X_n)$  para extrairmos toda informação que a amostra  $X_1, \dots, X_n$  contém sobre  $\theta$ , então dizemos que  $T$  (que pode ser um vetor) é suficiente para  $\theta$ . Desse modo, o conhecimento apenas de  $T$  (e não necessariamente da amostra completa  $X_1, \dots, X_n$ ) é suficiente para que sejam feitas inferências sobre  $\theta$ . A seguir apresentamos a definição formal.

**Definição 2.2.1.** Dizemos que a estatística  $T = T(X_1, \dots, X_n)$  é suficiente para  $\theta$ , quando a distribuição condicional de  $X_1, \dots, X_n$  dado  $T$  for independente de  $\theta$ .

Os exemplos a seguir ilustram a obtenção de estatísticas suficientes pela utilização da Definição 2.2.1.

**Exemplo 2.2.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição  $Binomial(1, \theta)$ , ou seja, de  $Bernoulli(\theta)$ . Vamos verificar se a estatística  $T = \sum_{i=1}^n X_i$  é suficiente para  $\theta$ . De acordo com a Definição 2.2.1,  $T$  é suficiente para  $\theta$ , se a probabilidade condicional  $P[X_1 = x_1, \dots, X_n = x_n | T = t]$  for independente de  $\theta$ . Temos, para  $x_1, \dots, x_n = 0$  ou  $1$  e  $t = 0, \dots, n$ ,

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

ou seja, sendo  $\sum_{i=1}^n x_i = t$ , temos que

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

pois sabemos que  $T \sim Binomial(n, \theta)$ . Portanto

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

de modo que, pela Definição 2.2.1,  $T = \sum_{i=1}^n X_i$  é suficiente para  $\theta$ .

**Exemplo 2.2.2.** Consideremos novamente a situação do Exemplo 2.2.1, com  $n = 3$  e  $T = X_1 + 2X_2 + X_3$ . Vamos verificar se  $T$  é suficiente. Notemos que para  $X_1 = 1, X_2 = 0, X_3 = 1$ , temos que  $T = 2$ . Logo

$$\begin{aligned} (2.2.1) \quad P[X_1 = 1, X_2 = 0, X_3 = 1 | T = 2] &= \frac{P[X_1 = 1, X_2 = 0, X_3 = 1]}{P[X_1 + 2X_2 + X_3 = 2]} \\ &= \frac{P[X_1 = 1]P[X_2 = 0]P[X_3 = 1]}{P[X_1 = 1, X_2 = 0, X_3 = 1] + P[X_1 = 0, X_2 = 1, X_3 = 0]} \\ &= \frac{\theta^2(1 - \theta)}{\theta^2(1 - \theta) + (1 - \theta)^2\theta} = \theta. \end{aligned}$$

Portanto, como a probabilidade (2.2.1) depende de  $\theta$ , concluímos que  $T$  não é suficiente para  $\theta$ , pois, nesse caso, a distribuição condicional de  $X_1, \dots, X_n$  dado  $T$  depende de  $\theta$ .

**Exemplo 2.2.3.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição de Poisson com parâmetro  $\theta$ . Vamos verificar se  $T = \sum_{i=1}^n X_i$  é suficiente para  $\theta$ . Sabemos que  $T = \sum_{i=1}^n X_i$  tem distribuição de Poisson com parâmetro  $n\theta$ . Assim, para  $x_i = 0, 1, 2, \dots, i = 1, \dots, n$  e  $t = 0, 1, \dots$ , temos

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

de modo que se  $\sum_{i=1}^n x_i = t$ , então,

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

que é independente de  $\theta$ . Segue, então, da Definição 2.2.1 que  $\sum_{i=1}^n X_i$  é suficiente para  $\theta$ .

Notemos que a Definição 2.2.1 permite, apenas, que possamos verificar se determinada estatística é ou não suficiente. Contudo não pode ser utilizada como um método para obtenção de estatísticas suficientes. Um procedimento para a obtenção de estatísticas suficientes é o critério da fatoração que apresentamos a seguir.

**Teorema 2.2.1.** (*Critério da Fatoração de Neyman*) Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição da variável aleatória  $X$  com função de densidade (ou de probabilidade)  $f(x|\theta)$  e função de verossimilhança  $L(\theta; \mathbf{x})$ . Temos, então, que a estatística  $T = T(X_1, \dots, X_n)$  é suficiente para  $\theta$ , se e somente se pudermos escrever

$$(2.2.2) \quad L(\theta; \mathbf{x}) = h(x_1, \dots, x_n)g_\theta(T(x_1, \dots, x_n)),$$

onde  $h(x_1, \dots, x_n)$  é uma função que depende apenas de  $x_1, \dots, x_n$  (não depende de  $\theta$ ) e  $g_\theta(T(x_1, \dots, x_n))$  depende de  $\theta$  e de  $x_1, \dots, x_n$  somente através de  $T$ .

**Prova.** Vamos provar o teorema apenas para o caso discreto. Nesse caso,  $L(\theta; \mathbf{x}) = P_\theta[\mathbf{X} = \mathbf{x}]$ . Suponhamos em primeiro lugar que (2.2.2) esteja verificada e então,

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

Como

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

quando  $T(\mathbf{x}) \neq t$ ,  $P[\mathbf{X} = \mathbf{x}|T(\mathbf{x}) = t] = 0$ , que é independente de  $\theta$ , logo a condição da Definição 2.2.1 está verificada. Quando  $T(\mathbf{x}) = t$ , o evento  $\{\mathbf{X} = \mathbf{x}, T(\mathbf{X}) = t\}$  está contido no evento  $\{T(\mathbf{x}) = t\}$ , então

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

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

Suponhamos agora que  $T = T(\mathbf{X})$  seja suficiente, de modo que a distribuição condicional de  $\mathbf{X}$  dado  $T$  é independente de  $\theta$ . Sendo  $T(\mathbf{x}) = t$ , temos que

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

de modo que (2.2.2) está provada.

**Exemplo 2.2.4.** Consideremos novamente o modelo de Poisson do Exemplo 2.2.3. Temos, então, que

$$\begin{aligned} L(\theta; \mathbf{x}) &= \prod_{i=1}^n f(x_i|\theta) \\ &= \frac{e^{-\theta}\theta^{x_1}}{x_1!} \cdots \frac{e^{-\theta}\theta^{x_n}}{x_n!} = \frac{1}{x_1! \cdots x_n!} e^{-n\theta} \theta^{x_1 + \dots + x_n}. \end{aligned}$$

Portanto, tomando

$$h(x_1, \dots, x_n) = \frac{1}{\prod_{i=1}^n x_i!} \prod_{i=1}^n I_{\{0,1,2,\dots\}}(x_i) \quad \text{e} \quad g_\theta(T(\mathbf{x})) = e^{-n\theta} \theta^{\sum_{i=1}^n x_i},$$

temos, pelo critério da fatoração, que  $T(\mathbf{X}) = \sum_{i=1}^n X_i$  é suficiente para  $\theta$ , onde  $\mathbf{X} = (X_1, \dots, X_n)$ .

**Exemplo 2.2.5.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim U(0, \theta)$ . Conforme visto no Capítulo 1, temos que (veja o Modelo 1.1.5)

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

Temos então

$$\begin{aligned} L(\theta; \mathbf{x}) &= \frac{1}{\theta} I_{[0,\theta]}(x_1) \dots \frac{1}{\theta} I_{[0,\theta]}(x_n) \\ &= \frac{1}{\theta^n} I_{[0,\theta]}(x_{(n)}) I_{[0,x_{(n)}]}(x_{(1)}), \end{aligned}$$

de modo que, pelo critério da fatoração,  $X_{(n)}$  é uma estatística suficiente para  $\theta$ .

**Exemplo 2.2.6.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição  $N(\mu, 1)$ . Temos, então, que

$$\begin{aligned} L(\mu; \mathbf{x}) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_1-\mu)^2}{2}} \dots \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_n-\mu)^2}{2}} \\ &= \left( \frac{1}{\sqrt{2\pi}} \right)^n e^{-\sum_{i=1}^n \frac{(x_i-\mu)^2}{2}} \\ &= \left( \frac{1}{\sqrt{2\pi}} \right)^n e^{-\sum_{i=1}^n \frac{x_i^2}{2}} e^{-\frac{n\mu^2}{2} + \mu \sum_{i=1}^n x_i}. \end{aligned}$$

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

## 2.3 Estatísticas Conjuntamente Suficientes

Na seção anterior vimos o caso uniparamétrico, ou seja, a distribuição dos dados depende de um único parâmetro  $\theta$ . Nesta seção consideramos o caso multiparamétrico em que  $\theta$  é um vetor de parâmetros, que denotamos por  $\theta$ . Em muitas situações, o modelo estatístico depende de mais de um parâmetro. É o caso do modelo  $N(\mu, \sigma^2)$ , em que  $\theta = (\mu, \sigma^2)$ , sendo  $\mu$  e  $\sigma^2$  desconhecidos.

É o caso também do modelo  $Gama(\alpha, \beta)$ , em que  $\alpha$  e  $\beta$  são desconhecidos e, portanto,  $\theta = (\alpha, \beta)$ .

**Teorema 2.3.1.** (*Critério da fatoração. Caso Multiparamétrico*) Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição da variável aleatória  $X$ , com função de densidade (ou de probabilidade)  $f(x|\theta)$ . Temos, então, que a estatística  $r$ -dimensional  $\mathbf{T} = (T_1, \dots, T_r)$ ,  $T_i = T_i(\mathbf{X})$  é conjuntamente suficiente para  $\theta$  se

$$L(\theta; \mathbf{x}) = f(x_1, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta) = h(x_1, \dots, x_n)g_\theta(T_1(\mathbf{x}), \dots, T_r(\mathbf{x})),$$

onde  $h(x_1, \dots, x_n)$  é uma função que não depende de  $\theta$  e  $g_\theta(T_1(\mathbf{x}), \dots, T_r(\mathbf{x}))$  depende de  $\theta$  e de  $\mathbf{x} = (x_1, \dots, x_n)$  somente por meio de  $(T_1(\mathbf{x}), \dots, T_r(\mathbf{x}))$ .

No caso do Teorema 2.3.1, dizemos que a estatística suficiente é de dimensão  $r$ , que em muitos casos é também a dimensão do espaço paramétrico  $\Theta$ . Mas existem situações em que tal fato não ocorre, ou seja, a dimensão de  $\Theta$  é menor que  $r$ .

**Exemplo 2.3.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X \sim N(\mu, \sigma^2)$ , onde  $\mu$  e  $\sigma^2$  são desconhecidos. Temos, então, que  $\theta = (\mu, \sigma^2)$ . Nesse caso, a função de verossimilhança pode ser escrita como

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

com  $-\infty < \mu < \infty$  e  $\sigma^2 > 0$ . Tomando  $h(x_1, \dots, x_n) = 1/(\sqrt{2\pi})^n$  e

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

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

**Definição 2.3.1.** 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  $\theta$ . Esse resultado vale também para o caso multidimensional.

**Exemplo 2.3.2.** Consideremos novamente a situação do Exemplo 2.2.6. Vamos 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 2.3.3.** Consideremos novamente a situação do Exemplo 2.3.1. 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  $\theta$  (Exemplo 2.3.1), temos que  $\mathbf{T}_2$  também é suficiente para  $\theta = (\mu, \sigma^2)$ .

**Exemplo 2.3.4.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com distribuição  $Gama(\alpha, \beta)$ . Dizemos que  $X \sim Gama(\alpha, \beta)$ , se sua f.d.p. é dada por

$$f(x|\alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \quad x > 0, \quad \alpha, \beta > 0.$$

onde  $\Gamma(\cdot)$  é a função gama definida por  $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$ , para  $t > 0$ . Então,  $\theta = (\alpha, \beta)$ . Temos que a função de verossimilhança correspondente à amostra observada é dada por

$$L(\theta; \mathbf{x}) = \frac{\beta^{n\alpha}}{\Gamma^n(\alpha)} \prod_{i=1}^n x_i^{\alpha-1} e^{-\beta \sum_{i=1}^n x_i} I_{(0, \infty)}(\mathbf{x}),$$

$\alpha > 0, \beta > 0$ . Portanto, pelo critério da fatoração, temos que  $\mathbf{T}_1 = (\prod_{i=1}^n X_i, \sum_{i=1}^n X_i)$  é conjuntamente suficiente para  $\theta$ . Notemos que a estatística  $\mathbf{T}_2 = (\sum_{i=1}^n \log X_i, \bar{X})$  é equivalente a  $\mathbf{T}_1$ .

## 2.4 Famílias Exponenciais

Muitos dos modelos estatísticos considerados nas seções anteriores podem ser considerados como casos especiais de uma família mais geral de distribuições.

**Definição 2.4.1.** Dizemos que a distribuição da variável aleatória  $X$  pertence à família exponencial unidimensional de distribuições, se pudermos escrever sua f.p. ou f.d.p. como

$$(2.4.1) \quad f(x|\theta) = e^{c(\theta)T(x)+d(\theta)+S(x)}, \quad x \in A$$

onde  $c, d$  são funções reais de  $\theta$ ;  $T, S$  são funções reais de  $x$  e  $A$  não depende de  $\theta$ .

Notemos que no caso em que  $X$  é contínua, para que  $f(x|\theta)$  em (2.4.1) seja uma função de densidade, é necessário que

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

ou seja,

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

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 2.4.1.** Seja  $X$  uma variável aleatória com distribuição de Bernoulli( $\theta$ ). Então, podemos escrever

$$f(x|\theta) = \theta^x(1-\theta)^{1-x} = \left(\frac{\theta}{1-\theta}\right)^x (1-\theta) = e^{x \log(\frac{\theta}{1-\theta}) + \log(1-\theta)}, \quad x = \{0, 1\}.$$

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 2.4.2.** Seja  $X$  uma variável aleatória com distribuição  $N(\mu, 1)$ . Temos, então, que

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

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

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

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

Outras distribuições que podem ser colocadas na forma da família exponencial unidimensional são, por exemplo, binomial, de Poisson e exponencial. O próximo resultado estabelece que amostras aleatórias de famílias exponenciais unidimensionais são também membros da família exponencial unidimensional.

**Teorema 2.4.1.** *Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$ , com função de densidade (ou de probabilidade) dada por (2.4.1). Então, a distribuição conjunta de  $X_1, \dots, X_n$  é dada por*

$$(2.4.2) \quad f(x_1, \dots, x_n|\theta) = e^{c^*(\theta) \sum_{i=1}^n T(x_i) + d^*(\theta) + S^*(\mathbf{x})}, \quad \mathbf{x} \in A^n,$$

que também é da família exponencial com  $T(\mathbf{x}) = \sum_{i=1}^n T(x_i)$ ,  $c^*(\theta) = c(\theta)$ ,  $d^*(\theta) = nd(\theta)$ , e  $S^*(\mathbf{x}) = \sum_{i=1}^n S(x_i)$ .



Notemos de (2.4.2) que considerando

$$h(x_1, \dots, x_n) = e^{\sum_{i=1}^n S(x_i)} \prod_{i=1}^n I_A(x_i), \quad \text{e} \quad g_\theta(T) = e^{c(\theta) \sum_{i=1}^n T(x_i) + nd(\theta)},$$

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

**Definição 2.4.2.** Dizemos que a distribuição da variável aleatória (ou de um vetor aleatório)  $X$  pertence à família exponencial de dimensão  $k$  se a função de densidade (ou de probabilidade) de  $X$  é dada por

$$(2.4.3) \quad f(x|\theta) = e^{\sum_{j=1}^k c_j(\theta) T_j(x) + d(\theta) + S(x)}, \quad x \in A,$$

onde  $c_j$ ,  $T_j$ ,  $d$  e  $S$  são funções reais,  $j = 1, \dots, k$ , e como no caso unidimensional,  $d(\theta)$  está associado à constante de normalização de (2.4.3) e  $A$  não depende de  $\theta$ .

Também, no caso de dimensão  $k$ , amostras de famílias exponenciais de dimensão  $k$  têm distribuições que são membros da família exponencial de dimensão  $k$ . Para uma amostra  $X_1, \dots, X_n$  de uma variável aleatória com função de densidade (ou de probabilidade) dada por (2.4.3), temos que a função de densidade (ou de probabilidade) conjunta de  $X_1, \dots, X_n$  é dada por

$$f(x_1, \dots, x_n|\theta) = e^{\sum_{j=1}^k c_j^*(\theta) \sum_{i=1}^n T_j(\mathbf{x}_i) + d^*(\theta) + S^*(\mathbf{x})},$$

onde

$$T_j^*(\mathbf{x}) = \sum_{i=1}^n T_j(x_i), \quad c_j^*(\theta) = c_j(\theta),$$

$$S^*(\mathbf{x}) = \sum_{i=1}^n S(x_i), \quad d^*(\theta) = nd(\theta).$$

Nesse caso,  $(T_1^*, \dots, T_k^*)$  é conjuntamente suficiente para  $\theta$ .

**Exemplo 2.4.3.** Consideremos mais uma vez a situação do Exemplo 2.3.1. Nesse caso, temos que  $\theta = (\mu, \sigma^2)$ , com

$$(2.4.4) \quad f(x|\theta) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

$$= e^{-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log\sigma^2 - \log\sqrt{2\pi}},$$

que é da família exponencial bidimensional com

$$T_1(x) = x, \quad T_2(x) = x^2, \quad c_1(\theta) = \frac{\mu}{\sigma^2}, \quad c_2(\theta) = -\frac{1}{2\sigma^2},$$

$$d(\theta) = -\frac{\mu}{2\sigma^2} - \frac{1}{2} \log \sigma^2, \quad S(x) = -\log \sqrt{2\pi}, \quad A = \mathbb{R}.$$

A distribuição de uma amostra aleatória da densidade (2.4.4) é também da família exponencial com  $T_1(\mathbf{X}) = \sum_{i=1}^n X_i$  e  $T_2(\mathbf{X}) = \sum_{i=1}^n X_i^2$ , que são conjuntamente suficientes para  $(\mu, \sigma^2)$ .

**Exemplo 2.4.4.** Vamos considerar agora o caso em que o vetor  $(X, Y)$  é distribuído de acordo com a distribuição normal bivariada com  $\theta = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ , que denotamos por

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N_2 \left( \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}; \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \right),$$

e com densidade

$$(2.4.5) \quad f(x, y|\theta) = \frac{\sigma_x^{-1}\sigma_y^{-1}}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)} \left[ \frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho}{\sigma_x\sigma_y} (x-\mu_x)(y-\mu_y) + \frac{(y-\mu_y)^2}{\sigma_y^2} \right]}.$$

A densidade pode ser escrita como

$$\begin{aligned} f(x, y|\theta) = & e^{\frac{1}{(1-\rho^2)} \left[ \frac{\mu_x}{\sigma_x^2} - \frac{\rho\mu_y}{\sigma_x\sigma_y} \right] x + \frac{1}{(1-\rho^2)} \left[ \frac{\mu_y}{\sigma_y^2} - \frac{\rho\mu_x}{\sigma_x\sigma_y} \right] y} \\ & e^{-\frac{1}{2(1-\rho^2)\sigma_x^2} x^2 - \frac{1}{2(1-\rho^2)\sigma_y^2} y^2 + \frac{\rho}{(1-\rho^2)\sigma_x\sigma_y} xy} \\ & e^{-\frac{\mu_x^2}{2(1-\rho^2)\sigma_x^2} - \frac{\mu_y^2}{2(1-\rho^2)\sigma_y^2} + \frac{\rho\mu_x\mu_y}{(1-\rho^2)\sigma_x\sigma_y} - \log \sigma_x\sigma_y \sqrt{1-\rho^2} - \log 2\pi}, \end{aligned}$$

que corresponde a uma densidade na forma da família exponencial de dimensão 5, em que

$$c_1(\theta) = \frac{1}{(1-\rho^2)} \left[ \frac{\mu_x}{\sigma_x^2} - \frac{\rho\mu_y}{\sigma_x\sigma_y} \right], \quad T_1(x, y) = x,$$

$$c_2(\theta) = \frac{1}{(1-\rho^2)} \left[ \frac{\mu_y}{\sigma_y^2} - \frac{\rho\mu_x}{\sigma_x\sigma_y} \right], \quad T_2(x, y) = y,$$

$$c_3(\theta) = -\frac{1}{2(1-\rho^2)\sigma_x^2}, \quad T_3(x, y) = x^2,$$

$$c_4(\theta) = -\frac{1}{2(1-\rho^2)\sigma_y^2}, \quad T_4(x, y) = y^2,$$

$$c_5(\theta) = \frac{\rho}{(1-\rho^2)\sigma_x\sigma_y}, \quad T_5(x, y) = xy.$$

As funções  $d(\theta)$  e  $S(x, y)$  são obtidas de maneira similar.

Consideremos uma amostra aleatória  $(X_1, Y_1), \dots, (X_n, Y_n)$  da densidade normal bivariada (2.4.5). Temos, portanto, que a estatística

$$\mathbf{T}_1 = \left( \sum_{i=1}^n X_i, \sum_{i=1}^n Y_i, \sum_{i=1}^n X_i^2, \sum_{i=1}^n Y_i^2, \sum_{i=1}^n X_i Y_i \right)$$

é conjuntamente suficiente para  $\theta = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ . Notemos que a estatística

$$\mathbf{T}_2 = (\bar{X}, \bar{Y}, S_x^2, S_y^2, S_{xy}),$$

onde  $S_x^2 = \sum_{i=1}^n (X_i - \bar{X})^2/n$ ,  $S_y^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2/n$  e  $S_{xy} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})/n$  é equivalente a  $\mathbf{T}_1$  e, portanto, é também conjuntamente suficiente para  $\theta$ . Estimadores comumente considerados para  $\theta$  e que são funções de  $\mathbf{T}_2$  são

$$(2.4.6) \quad \hat{\mu}_x = \bar{X}, \quad \hat{\mu}_y = \bar{Y}, \quad \hat{\sigma}_x^2 = \sum_{i=1}^n (X_i - \bar{X})^2/n, \quad \hat{\sigma}_y^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2/n,$$

e

$$(2.4.7) \quad \hat{\rho} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}.$$

O estimador  $\hat{\rho}$  é conhecido como coeficiente de correlação de Pearson. Podemos mostrar que os estimadores de  $\theta$  dados por (2.4.6) e (2.4.7) são estimadores de máxima verossimilhança.

## 2.5 Estimadores Baseados em Estatísticas Suficientes

Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com função de densidade (ou de probabilidade)  $f(x|\theta)$ . Seja  $T = T(X_1, \dots, X_n)$  uma estatística suficiente para  $\theta$  e  $S = S(X_1, \dots, X_n)$  um estimador de  $\theta$  que não é função da estatística suficiente  $T$ . Então,

$$(2.5.1) \quad \hat{\theta} = E[S|T],$$

é um estimador de  $\theta$ , ou seja, é uma função de  $T$  que não depende de  $\theta$ , pois, sendo  $T$  suficiente, a distribuição condicional de  $X_1, \dots, X_n$  dado  $T$  é independente de  $\theta$ . Notemos que  $S = S(X_1, \dots, X_n)$  é uma função apenas de  $X_1, \dots, X_n$ . Temos, também, que se  $S$  é um estimador não viciado de  $\theta$ , então  $\hat{\theta}$  é também não viciado para  $\theta$  (veja o Exercício 2.8). Contudo o resultado mais

importante, conhecido como Teorema de Rao-Blackwell, estabelece que, se  $S$  é um estimador não viciado de  $\theta$ , então,

$$(2.5.2) \quad \text{Var}[\hat{\theta}] \leq \text{Var}[S],$$

para todo  $\theta$ . Para provar esse resultado, notemos que

$$\begin{aligned} \text{Var}[S] &= E\{\text{Var}[S|T]\} + \text{Var}\{E[S|T]\} \\ &\geq \text{Var}\{E[S|T]\} = \text{Var}[\hat{\theta}], \end{aligned}$$

pois  $E\{\text{Var}[S|T]\} \geq 0$ . Portanto temos de (2.5.2) que o estimador  $\hat{\theta}$  baseado na estatística suficiente  $T$  apresenta uma variância menor (ou igual) que a variância do estimador não viciado  $S$ . Desse modo, qualquer estimador  $S$  que não é função de uma estatística suficiente pode ser melhorado pelo procedimento (2.5.1).

**Exemplo 2.5.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim \text{Poisson}(\theta)$ . Queremos estimar  $P(X = 0) = \tau = e^{-\theta}$ . Temos que a estatística  $T = \sum_{i=1}^n X_i$  é suficiente para  $\theta$ . Consideremos

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

Temos que  $E(S) = P(X_1 = 0) = e^{-\theta}$ , logo  $S$  é não viciado para  $e^{-\theta}$ . 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 de acordo com (2.5.1) temos que o estimador

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

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

A seguir apresentamos a definição de estatística completa que, em conjunto com a definição de suficiência, possibilita a obtenção do estimador ótimo, isto é, o estimador não viciado de variância uniformemente mínima.

**Definição 2.5.1.** Uma estatística  $T = T(X_1, \dots, X_n)$  é dita ser completa em relação à família  $f(x|\theta) : \theta \in \Theta$ , se a única função real  $g$ , definida no domínio

de  $T$ , tal que  $E[g(T)] = 0$ , para todo  $\theta$  é a função nula, isto é,  $g(T) = 0$  com probabilidade 1.

**Exemplo 2.5.2.** Consideremos novamente o Exemplo 2.2.1. Temos que

$$E[g(T)] = \sum_{x=0}^n g(x) \binom{n}{x} \theta^x (1-\theta)^{n-x} = 0 \quad \text{para todo } \theta,$$

se e somente se

$$(2.5.3) \quad \sum_{x=0}^n g(x) \binom{n}{x} \rho^x = 0, \quad \text{para todo } \rho$$

onde  $\rho = \theta/(1-\theta)$ . Como o lado esquerdo de (2.5.3) é um polinômio em  $\rho$  de grau  $n$  temos que  $g(x) = 0$  para todo  $x$ . Portanto  $T = \sum_{i=1}^n X_i$  é completa em relação à família Binomial.

**Exemplo 2.5.3.** Sejam  $X_1, X_2$  uma amostra aleatória da variável  $X \sim \text{Bernoulli}(\theta)$ . Seja  $T = X_1 - X_2$ . Temos que  $E(T) = E(X_1 - X_2) = 0$ , logo existe a função  $g(T) = T$  tal que  $E(g(T)) = 0$ , mas  $g(T) \neq 0$  com probabilidade 1. Portanto  $T = X_1 - X_2$  não é completa.

A demonstração do teorema a seguir pode ser encontrada em Lehmann (1986).

**Teorema 2.5.2.** *Suponha que  $X$  tenha distribuição da família exponencial  $k$ -dimensional (como definida em 2.4.2). Então, a estatística*

$$T(\mathbf{X}) = \left( \sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i) \right)$$

*é suficiente para  $\theta$ .  $T(\mathbf{X})$  será também completa desde que o domínio de variação de  $(c_1(\theta), \dots, c_k(\theta))$  contenha um retângulo  $k$ -dimensional.*

No caso uniparamétrico, é necessário que o domínio de variação de  $c(\theta)$  contenha um intervalo da reta. No caso bidimensional, um quadrado e assim por diante.

**Teorema 2.5.3.** (Lehmann-Scheffé) *Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com f.d.p. (ou f.p.),  $f(x|\theta)$ . Seja  $T$  uma estatística suficiente e completa. Seja  $S$  um estimador não viciado de  $\theta$ . Então  $\hat{\theta} = E(S|T)$  é o único estimador não viciado de  $\theta$  baseado em  $T$  e é o estimador não viciado de variância uniformemente mínima (ENVVUM) para  $\theta$ .*

**Prova.** De (2.5.1) e (2.5.2) temos que  $\hat{\theta}$  é um estimador não viciado de  $\theta$  e que, na procura de ENVVUM's para  $\theta$ , basta procurar entre os que são função de

$T$  (pois os que não são podem ser melhorados). Falta provar, então, que há um único estimador não viciado de  $\theta$  que é função de  $T$ . Para isso, suponha que existam  $\hat{\theta}_1$  e  $\hat{\theta}_2$ , ambos funções de  $T$ , tais que

$$E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta,$$

de modo que  $E(\hat{\theta}_1 - \hat{\theta}_2) = 0$  e como  $T$  é completa,  $\hat{\theta}_1 - \hat{\theta}_2 = 0$ , e portanto  $\hat{\theta}_1 = \hat{\theta}_2$  com probabilidade 1.

**Exemplo 2.5.4.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição de Poisson com parâmetro  $\theta$ . Pelos Exemplos 2.2.4 e 2.5.2 temos que  $T = \sum_{i=1}^n X_i$  é uma estatística suficiente e completa. Como  $\bar{X}$  é um estimador não viciado de  $\theta$  e é função de  $T$ , é o ENVVUM.

## 2.6 Exercícios

**2.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(0, \sigma^2)$ .

- (i) Encontre o limite inferior da variância dos estimadores não viciados de  $\sigma^2$ .
- (ii) Encontre uma estatística suficiente para  $\sigma^2$ .
- (iii) Obtenha a partir desta estatística um estimador não viciado para  $\sigma^2$ .
- (iv) Verifique se este estimador é eficiente.

**2.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim \text{Binomial}(2, \theta)$ .

- (i) Encontre o limite inferior da variância dos estimadores não viciados de  $\theta$ .
- (ii) Encontre uma estatística suficiente para  $\theta$ .
- (iii) Obtenha um estimador não viciado para  $\theta$  que seja função da estatística suficiente.
- (iv) Verifique se o estimador é eficiente.

**2.3.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição da variável aleatória  $X$  com função densidade de probabilidade dada por

$$f(x|\theta) = \theta x^{\theta-1}, \quad 0 < x < 1, \quad \theta > 0.$$

- (i) Mostre que a f.d.p. pertence à família exponencial.
- (ii) Encontre o limite inferior da variância dos estimadores não viciados de  $\theta$ .
- (iii) Encontre uma estatística suficiente para  $\theta$  e sua distribuição.
- (iv) Sugira um estimador não viciado para  $\theta$  que seja função da estatística suficiente e verifique se é eficiente.

**2.4.** Sejam  $X_1, X_2$  uma amostra aleatória da variável aleatória  $X \sim \text{Poisson}(\theta)$ . Mostre que  $T = X_1 + 2X_2$  não é suficiente para  $\theta$ .

**2.5.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com função de densidade (ou de probabilidade)  $f(x|\theta)$  para a qual as condições de regularidade estão satisfeitas. Seja  $\hat{\gamma}$  um estimador não viciado para  $g(\theta)$ . Mostre que

$$\text{Var}(\hat{\gamma}) \geq \frac{(g'(\theta))^2}{nE \left[ \left( \frac{\partial \log f(X|\theta)}{\partial \theta} \right)^2 \right]}.$$

**2.6.** Seja  $f(x|\theta)$  uma função densidade para a qual as condições de regularidade estão satisfeitas. Mostre que

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

**2.7.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com f.d.p. dada por

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

- (i) Encontre uma estatística suficiente para  $\theta$ .
- (ii) Baseado nesta estatística, obtenha um estimador não viciado para  $\theta$ .

**2.8.** Mostre que se  $S$  é um estimador não viciado de  $\theta$ , então  $\hat{\theta}$  dado por (2.5.1) também é não viciado para  $\theta$ .

**2.9.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\mu, 1)$ .

- (i) Mostre que o estimador  $\hat{\gamma} = \bar{X}^2 - 1/n$  é não viciado para  $g(\mu) = \mu^2$ .
- (ii) Existe ENVVUM para  $\mu^2$ ?
- (iii) Encontre o limite inferior da variância dos estimadores não viciados de  $g(\mu) = \mu^2$  e verifique se  $\hat{\gamma}$  é eficiente.

**2.10.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim \text{Bernoulli}(\theta)$ . Obtenha o ENVVUM para  $\theta(1 - \theta)$ .

Sugestão: verifique se  $S^2 = \frac{n}{(n-1)}\bar{X}(1 - \bar{X})$  é não viciado para  $\theta(1 - \theta)$ .

**2.11.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com distribuição geométrica com parâmetro  $\theta$ , isto é,

$$f(x|\theta) = \theta(1 - \theta)^x, \quad x = 0, 1, 2, \dots, \quad 0 < \theta < 1.$$

Encontre o ENVVUM para  $1/\theta$ .

**2.12.** Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias independentes onde  $Y_i \sim N(\beta x_i, \sigma^2)$ , onde  $x_i$  é conhecido,  $i = 1, \dots, n$ . Note que, neste caso, as variáveis  $Y_i$  não são identicamente distribuídas.

- (i) Encontre uma estatística conjuntamente suficiente para  $\beta$  e  $\sigma^2$ .
- (ii) Baseado nessa estatística, obtenha os ENVVUM para  $\beta$  e para  $\sigma^2$ .



### 3. Métodos de Estimação

No capítulo anterior consideramos um critério para verificar se determinado estimador é ou não eficiente. Contudo tal procedimento não é um método que possibilita, em geral, a obtenção de estimadores em situações específicas. Vimos também que todo bom estimador deve ser função de uma estatística suficiente. Neste capítulo vamos considerar alguns métodos que possibilitam a obtenção de estimadores em situações específicas. O primeiro método que consideramos é o método de máxima verossimilhança em que estimadores são obtidos a partir da maximização da função de verossimilhança. O segundo método considerado é o método dos momentos em que estimadores são obtidos igualando-se os momentos amostrais aos correspondentes momentos populacionais.

#### 3.1 O Método de Máxima Verossimilhança

O conceito de função de verossimilhança, enunciado a seguir, é central na teoria da verossimilhança.

**Definição 3.1.1.** *Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$  com função de densidade (ou de probabilidade)  $f(x|\theta)$ , com  $\theta \in \Theta$ , onde  $\Theta$  é o espaço paramétrico. A função de verossimilhança de  $\theta$  correspondente à amostra aleatória observada é dada por*

$$(3.1.1) \quad L(\theta; \mathbf{x}) = \prod_{i=1}^n f(x_i|\theta).$$

**Definição 3.1.2.** *O estimador de máxima verossimilhança de  $\theta$  é o valor  $\hat{\theta} \in \Theta$  que maximiza a função de verossimilhança  $L(\theta; \mathbf{x})$ .*

O logaritmo natural da função de verossimilhança de  $\theta$  é denotado por

$$(3.1.2) \quad l(\theta; \mathbf{x}) = \log L(\theta; \mathbf{x}).$$

Não é difícil verificar que o valor de  $\theta$  que maximiza a função de verossimilhança  $L(\theta; \mathbf{x})$ , também maximiza  $l(\theta; \mathbf{x})$  dada por (3.1.2). Além disso, no caso

uniparamétrico onde  $\Theta$  é um intervalo da reta e  $l(\theta; \mathbf{x})$  é derivável, o estimador de máxima verossimilhança pode ser encontrado como a raiz da equação de verossimilhança

$$(3.1.3) \quad l'(\theta; \mathbf{x}) = \frac{\partial l(\theta; \mathbf{x})}{\partial \theta} = 0.$$

Em alguns exemplos simples, a solução da equação de verossimilhança pode ser obtida explicitamente. Em situações mais complicadas, a solução da equação (3.1.3) será em geral obtida por procedimentos numéricos. Para se concluir que a solução da equação (3.1.3) é um ponto de máximo, é necessário verificar se

$$(3.1.4) \quad l''(\hat{\theta}; \mathbf{x}) = \frac{\partial^2 \log L(\theta; \mathbf{x})}{\partial \theta^2} \Big|_{\theta=\hat{\theta}} < 0.$$

Em situações em que  $\Theta$  é discreto ou em que o máximo de  $l(\theta; \mathbf{x})$  ocorre na fronteira de  $\Theta$  (Exemplo 1.3.8), o estimador de máxima verossimilhança não pode ser obtido a partir da solução de (3.1.3). Em tais situações, o máximo é obtido a partir da inspeção da função de verossimilhança.

**Exemplo 3.1.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição da variável aleatória  $X \sim N(\mu, 1)$ . Nesse caso, a função de verossimilhança é dada por

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

com  $\Theta = \{\mu; -\infty < \mu < \infty\}$ . Como

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

segue de (3.1.3) 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 nesse caso que (3.1.4) está satisfeita.

Então  $\bar{X}$ , além de ser eficiente (Exemplo 2.1.1) e função da estatística suficiente, é também estimador de máxima verossimilhança.

**Exemplo 3.1.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim \text{Bernoulli}(\theta)$ . Nesse caso, a função de verossimilhança de  $\theta$  é dada por

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

com  $\Theta = \{\theta; 0 < \theta < 1\}$ . De modo que

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

Portanto segue de (3.1.3) que a equação de verossimilhança de  $\theta$  é dada por

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

logo o estimador de máxima verossimilhança de  $\theta$  é dado por

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i,$$

pois neste caso, (3.1.4) também está verificada.

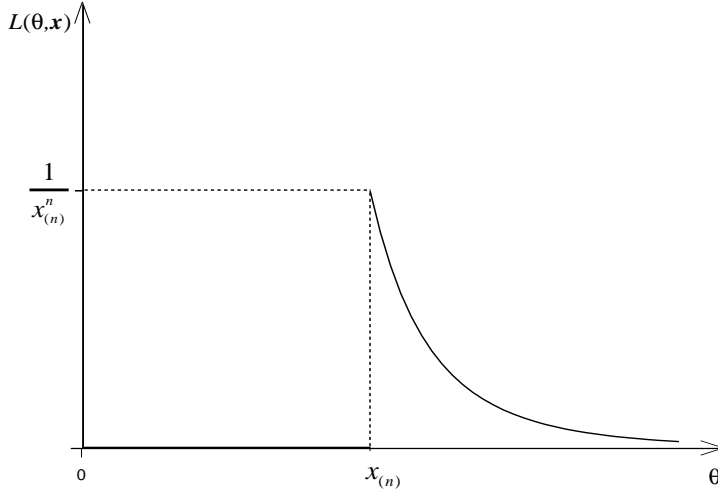
O exemplo a seguir ilustra uma situação em que a equação (3.1.3) não pode ser utilizada.

**Exemplo 3.1.3.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim U(0, \theta)$ . Conforme visto no Exemplo 2.2.5, podemos escrever a função de verossimilhança como

$$(3.1.5) \quad L(\theta; \mathbf{x}) = \frac{1}{\theta^n} I_{[0, \theta]}(x_{(n)}) I_{[0, x_{(n)}]}(x_{(1)}),$$

onde  $\Theta = \{\theta; \theta > 0\}$ . Nesse caso, a equação de verossimilhança (3.1.3) não leva a nenhum estimador para  $\theta$ . Por outro lado, o gráfico da função de verossimilhança de  $\theta$  é dado pela Figura 3.1.

Como a função de verossimilhança (3.1.5) é nula para  $\theta < x_{(n)}$  e vale  $1/\theta^n$  para  $\theta \geq x_{(n)}$ , temos que o máximo de  $L(\theta; \mathbf{x})$  é dado por  $\hat{\theta} = x_{(n)}$ , que é uma estatística suficiente para  $\theta$ . Nesse caso o estimador de máxima verossimilhança de  $\theta$  é viciado (ver Exemplo 1.3.8.).

**Figura 3.1.** Função de Verossimilhança

No caso discreto, o estimador de máxima verossimilhança de  $\theta$ ,  $\hat{\theta}$ , pode ser interpretado como o valor de  $\theta$  que maximiza a probabilidade de se observar a amostra que foi selecionada. O exemplo a seguir ilustra bem esse fato.

**Exemplo 3.1.4.** Temos uma caixa com bolas brancas e vermelhas. Sabe-se que a proporção  $\theta$  de bolas vermelhas na caixa é  $1/3$  ou  $2/3$ . Portanto  $\Theta = \{1/3, 2/3\}$ . Para obtermos informação sobre  $\theta$ , uma amostra de  $n = 3$  bolas é observada com reposição e apresenta bola vermelha na primeira extração e branca na segunda e na terceira extrações. Definindo

$$X_i = \begin{cases} 1, & \text{se a } i\text{-ésima retirada apresenta bola vermelha} \\ 0, & \text{se a } i\text{-ésima retirada apresenta bola branca,} \end{cases}$$

para  $i = 1, 2, 3$ , temos que a função de verossimilhança de  $\theta$  associada à amostra observada é dada por

$$L(\theta; \mathbf{x}) = P_\theta[X_1 = 1, X_2 = 0, X_3 = 0] = \theta(1 - \theta)(1 - \theta) = \theta(1 - \theta)^2.$$

Como

$$L\left(\frac{1}{3}; \mathbf{x}\right) = \frac{1}{3} \left(\frac{2}{3}\right)^2 = \frac{4}{27}$$

e

$$L\left(\frac{2}{3}; \mathbf{x}\right) = \frac{2}{3} \left(\frac{1}{3}\right)^2 = \frac{2}{27},$$

temos que a estimativa de máxima verossimilhança de  $\theta$  é dada por  $\hat{\theta} = 1/3$ , pois

$$L\left(\frac{1}{3}; \mathbf{x}\right) > L\left(\frac{2}{3}; \mathbf{x}\right).$$

O exemplo que apresentamos a seguir ilustra uma situação em que o estimador de máxima verossimilhança não é único.

**Exemplo 3.1.5.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição da variável aleatória  $X \sim U(\theta - 1/2, \theta + 1/2)$ , isto é

$$f(x|\theta) = I_{[\theta-1/2; \theta+1/2]}(x),$$

$\theta > 0$ . Temos, então, que

$$\begin{aligned} L(\theta; \mathbf{x}) &= I_{[\theta-1/2; \theta+1/2]}(x_1) \cdots I_{[\theta-1/2; \theta+1/2]}(x_n) \\ &= I_{[x_{(n)}-1/2; x_{(1)}+1/2]}(\theta), \end{aligned}$$

pois

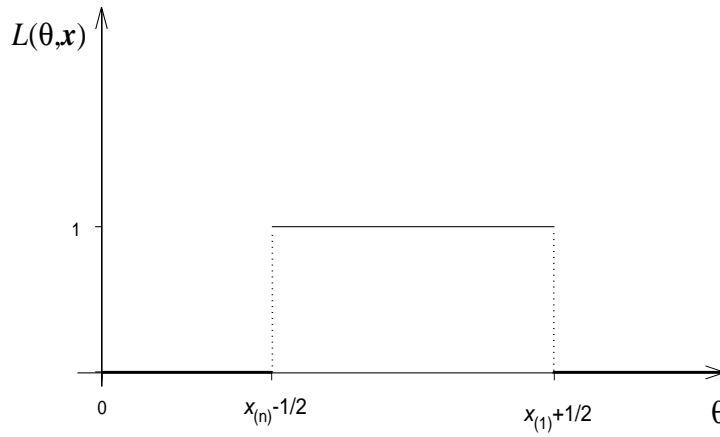
$$\theta - 1/2 \leq x_i \leq \theta + 1/2, \quad i = 1, \dots, n,$$

se e somente se

$$\theta \leq x_{(1)} + 1/2 \quad \text{e} \quad x_{(n)} - 1/2 \leq \theta.$$

A Figura 3.2 apresenta o gráfico da função  $L(\theta; \mathbf{x})$ .

**Figura 3.2.** Função de Verossimilhança



Como  $L(\theta; \mathbf{x})$  é nula para  $\theta < x_{(n)} - 1/2$  e para  $\theta > x_{(1)} + 1/2$  e constante no intervalo  $[x_{(n)} - 1/2; x_{(1)} + 1/2]$ , temos que qualquer ponto desse intervalo é um estimador de máxima verossimilhança de  $\theta$ . Em particular,

$$\hat{\theta} = \frac{X_{(1)} + X_{(n)}}{2}$$

é um estimador de máxima verossimilhança de  $\theta$ .

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. Vamos denotar por  $U(\theta)$  a função escore, ou seja,

$$U(\theta) = \frac{\partial \log L(\theta; \mathbf{x})}{\partial \theta},$$

temos que, para o estimador de máxima verossimilhança  $\hat{\theta}$ ,

$$U(\hat{\theta}) = 0,$$

de modo que, expandindo  $U(\hat{\theta})$  em série de Taylor em torno de um ponto  $\theta_0$ , obtemos

$$0 = U(\hat{\theta}) \cong U(\theta_0) + (\hat{\theta} - \theta_0)U'(\theta_0),$$

ou seja, chegamos a equação

$$(3.1.6) \quad \hat{\theta} \cong \theta_0 - \frac{U(\theta_0)}{U'(\theta_0)}.$$

Da equação (3.1.6), obtemos o procedimento iterativo (Newton-Raphson)

$$(3.1.7) \quad \theta_{j+1} = \theta_j - \frac{U(\theta_j)}{U'(\theta_j)},$$

que é iniciado com o valor  $\theta_0$  e então um novo valor  $\theta_1$  é obtido a partir de (3.1.7) e assim por diante, até que o processo se estabilize, ou seja, para um dado  $\epsilon$  pequeno,  $|\theta_{j+1} - \theta_j| < \epsilon$ . Nesse caso, o ponto  $\hat{\theta}$  em que o processo se estabiliza é tomado como o estimador de máxima verossimilhança de  $\theta$ . Em alguns casos, a substituição de  $U'(\theta_j)$  em (3.1.7) por  $E[U'(\theta_j)]$ , ou seja, a informação de Fisher em  $\theta_j$  correspondente à amostra observada multiplicada por  $-1$ , apresenta significativa simplificação no procedimento. Esse método é conhecido como método do escore. O exemplo a seguir ilustra uma aplicação de tal procedimento.

**Exemplo 3.1.6.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição da variável aleatória  $X$  com função de densidade dada por

$$(3.1.8) \quad f(x|\theta) = \frac{1}{2}(1 + \theta x); \quad -1 \leq x \leq 1, \quad -1 \leq \theta \leq 1.$$

Nesse caso,

$$L(\theta; \mathbf{x}) = \frac{1}{2^n} \prod_{i=1}^n (1 + \theta x_i),$$

de modo que

$$U(\theta) = \frac{\partial \log L(\theta; \mathbf{x})}{\partial \theta} = \sum_{i=1}^n \frac{x_i}{1 + \theta x_i}.$$

Assim

$$U'(\theta) = - \sum_{i=1}^n \frac{x_i^2}{(1 + \theta x_i)^2},$$

de modo que o procedimento iterativo (3.1.7) se reduz a

$$(3.1.9) \quad \theta_{j+1} = \theta_j + \frac{\sum_{i=1}^n \frac{x_i}{1 + \theta_j x_i}}{\sum_{i=1}^n \frac{x_i^2}{(1 + \theta_j x_i)^2}}.$$

Podemos verificar que a informação de Fisher de  $\theta$  é dada, para  $\theta \neq 0$ , por

$$I_F(\theta) = \frac{1}{2\theta^3} \left\{ \log \left( \frac{1 + \theta}{1 - \theta} \right) - 2\theta \right\},$$

de modo que um procedimento alternativo a (3.1.9) é dado por

$$(3.1.10) \quad \theta_{j+1} = \theta_j - \frac{\sum_{i=1}^n \frac{x_i}{1 + \theta_j x_i}}{n I_F(\theta_j)}.$$

Uma amostra de tamanho  $n = 20$  é gerada a partir da densidade (3.1.8) com  $\theta = 0, 4$ . Os dados foram gerados a partir do método da função de distribuição, ou seja, sendo  $F(X) = U$ , temos que  $U \sim U(0, 1)$ . Nesse caso, como

$$F(x) = \int_{-1}^x \frac{1}{2}(1 + \theta y) dy = \frac{x + 1}{2} + \frac{\theta(x^2 - 1)}{4},$$

temos que se  $U \sim U(0, 1)$ , então,

$$(3.1.11) \quad x = \frac{-1 + 2\sqrt{1/4 - \theta(1/2 - \theta/4 - u)}}{\theta}$$

tem distribuição com função de densidade dada por (3.1.8), ou seja, para  $u$  gerado a partir da  $U(0, 1)$ ,  $x$  obtido a partir de (3.1.11) é um valor gerado a partir da distribuição com função de densidade dada por (3.1.8). As observações geradas são dadas na Tabela 3.1.

**Tabela 3.1.**  $n = 20$  observações da densidade (3.1.8) com  $\theta = 0,4$ 

0,3374	0,9285	0,6802	-0,2139	0,1052
-0,9793	-0,2623	-0,1964	0,5234	-0,0349
-0,6082	0,7509	0,3424	-0,7010	-0,2605
0,4077	-0,7435	0,9862	0,9704	0,5313

Escrevendo um programa em Fortran (outra linguagem poderia também ser facilmente utilizada) para calcular o estimador de máxima verossimilhança, obtemos, após 10 iterações do programa, a Tabela 3.2 em que a segunda coluna corresponde ao procedimento dado em (3.1.9) e a terceira coluna corresponde ao procedimento (3.1.10). Como valor inicial para o procedimento iterativo foi usado  $\theta_0 = \bar{X} = 0,1282$ .

**Tabela 3.2.** Valores de  $\hat{\theta}$  obtidos nas 10 iterações

Iteração	Usando (3.1.9)	Usando (3.1.10)
1	0,128188	0,128188
2	0,358745	0,371861
3	0,351170	0,349163
4	0,351140	0,351328
5	0,351140	0,351123
6	0,351140	0,351142
7	0,351140	0,351140
8	0,351140	0,351140
9	0,351140	0,351140
10	0,351140	0,351140

### 3.2 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 uma estatística suficiente.

**Teorema 3.2.1.** *Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com função de densidade (ou de probabilidade)  $f(x|\theta)$ . Seja  $T = T(X_1, \dots, X_n)$  uma estatística suficiente para  $\theta$ . Então o estimador de máxima verossimilhança  $\hat{\theta}$  (se existir) é função de  $T$ .*

**Prova.** De acordo com o critério da fatoração, temos que se  $T$  é suficiente para  $\theta$ , então,

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



onde  $g_\theta(T(\mathbf{x}))$  depende de  $\mathbf{x}$  somente através de  $T$ . Como  $h(\mathbf{x})$  é constante com relação a  $\theta$ , temos que maximizar  $L(\theta; \mathbf{x})$  com relação a  $\theta$  é equivalente a maximizar  $g_\theta(T(\mathbf{x}))$  com relação a  $\theta$ . Como  $g_\theta(T(\mathbf{x}))$  depende de  $\mathbf{x}$  somente através de  $T$ , temos que  $\hat{\theta}$  será obrigatoriamente uma função de  $T$ . Outras propriedades são apresentadas nas subseções seguintes.

### 3.2.1 Invariância

A seguir apresentamos uma propriedade bastante importante do método de máxima verossimilhança. Seja  $g(\cdot)$  uma função real 1 : 1 (invertível) definida em  $\Theta$ .

**Teorema 3.2.2.** *(O princípio da invariância.) Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com função de densidade (ou de probabilidade)  $f(x|\theta)$ . Se  $\hat{\theta}$  é um estimador de máxima verossimilhança de  $\theta$ , então  $g(\hat{\theta})$  é um estimador de máxima verossimilhança de  $g(\theta)$ .*

**Prova.** Provamos o resultado para o caso em que  $g$  é 1:1. Sendo  $g(\cdot)$  uma função 1 : 1, temos que  $g(\cdot)$  é invertível, de modo que  $\theta = g^{-1}(g(\theta))$ . Assim

$$(3.2.1) \quad L(\theta; \mathbf{x}) = L(g^{-1}(g(\theta)); \mathbf{x}),$$

de modo que  $\hat{\theta}$  maximiza os dois lados de (3.2.1). Logo

$$\hat{\theta} = g^{-1}(\widehat{g(\theta)}),$$

portanto

$$\widehat{g(\theta)} = g(\hat{\theta}),$$

ou seja, o estimador de máxima verossimilhança de  $g(\theta)$  é  $g(\hat{\theta})$ .

**Exemplo 3.2.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X \sim \text{Bernoulli}(\theta)$ . Nesse caso, o parâmetro de interesse é  $g(\theta) = \theta(1-\theta)$ . De acordo com o princípio da invariância, temos que o estimador de máxima verossimilhança de  $g(\theta)$  é dado por

$$(3.2.2) \quad g(\hat{\theta}) = \bar{X}(1 - \bar{X}).$$

De acordo com o Exercício 2.10 temos que o estimador dado em (3.2.2) é viciado para  $g(\theta)$ . Por outro lado, usando o Exercício 2.10, temos que

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

que decresce à medida que  $n$  aumenta.

**Exemplo 3.2.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição da variável aleatória  $X \sim N(\mu, 1)$ . Vimos que  $\hat{\mu} = \bar{X}$  é o estimador de máxima verossimilhança de  $\mu$ . Suponhamos que queremos estimar

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

Pelo princípio da invariância, temos que

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

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

**Exemplo 3.2.3.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição da variável aleatória  $X \sim \text{Exp}(\theta)$  com densidade

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

$\theta > 0$  e  $x > 0$ . Nesse caso,  $\hat{\theta} = \bar{X}^{-1}$  é o estimador de máxima verossimilhança de  $\theta$ . Suponhamos que é de interesse estimar

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

De acordo com o princípio da invariância, temos que o estimador de máxima verossimilhança é

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

Nos três exemplos acima, vimos situações em que o estimador de máxima verossimilhança é uma função complicada da amostra observada. 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 eficiente, em grandes amostras.

### 3.2.2 Distribuição em grandes amostras

No caso em que o tamanho da amostra é grande, e as condições de regularidade, especificadas no Capítulo 2, estão satisfeitas, temos que

$$(3.2.3) \quad \sqrt{n}(\hat{\theta} - \theta) \stackrel{a}{\sim} N\left(0, \frac{1}{I_F(\theta)}\right),$$

e

$$(3.2.4) \quad \sqrt{n}(g(\hat{\theta}) - g(\theta)) \stackrel{a}{\sim} N\left(0, \frac{(g'(\theta))^2}{I_F(\theta)}\right),$$

onde " $\stackrel{a}{\sim}$ " significa distribuição assintótica. Temos então que, para amostras grandes, os estimadores de máxima verossimilhança de  $\theta$  e  $g(\theta)$  são aproximadamente não viciados, cujas variâncias coincidem com os correspondentes limites inferiores das variâncias dos estimadores não viciados de  $\theta$  e  $g(\theta)$ . Portanto, em grandes amostras, temos que o estimador de máxima verossimilhança é eficiente.

**Exemplo 3.2.4.** Considere o modelo do Exemplo 3.2.1. De acordo com (3.2.4), temos que a distribuição do estimador de máxima verossimilhança (3.2.2) é dada por

$$\sqrt{n}(g(\hat{\theta}) - \theta(1 - \theta)) \stackrel{a}{\sim} N(0, (1 - 2\theta)^2 \theta(1 - \theta)),$$

pois  $g'(\theta) = 1 - 2\theta$ .

**Exemplo 3.2.5.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim \text{Poisson}(\theta)$ . Nesse caso, temos que o estimador de máxima verossimilhança de  $\theta$  é  $\hat{\theta} = \bar{X}$  (verifique!). De acordo com o princípio da invariância, temos que o estimador de máxima verossimilhança de  $e^{-\theta}$  é dado por

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

Do resultado (3.2.4), temos que

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

### 3.3 Verossimilhança para Amostras Independentes

Existem situações em que temos duas ou mais amostras independentes de distribuições que dependem de um parâmetro  $\theta$  de interesse. No caso de duas amostras aleatórias independentes,  $X_1, \dots, X_n$  e  $Y_1, \dots, Y_n$ , podemos escrever

$$(3.3.1) \quad L(\theta; \mathbf{x}, \mathbf{y}) = L(\theta; \mathbf{x})L(\theta; \mathbf{y}),$$

devido à independência entre as amostras. Portanto a verossimilhança conjunta é igual ao produto da verossimilhança correspondente à amostra  $X_1, \dots, X_n$  pela verossimilhança correspondente à amostra  $Y_1, \dots, Y_n$ . De (3.3.1), podemos escrever

$$l(\theta; \mathbf{x}, \mathbf{y}) = l(\theta; \mathbf{x}) + l(\theta; \mathbf{y}),$$

de modo que o logaritmo da verossimilhança conjunta é igual à soma do logaritmo das verossimilhanças correspondentes a cada uma das amostras. O exemplo que apresentamos a seguir ilustra uma tal situação.

**Exemplo 3.3.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória correspondente a  $X \sim N(\mu, 4)$  e  $Y_1, \dots, Y_n$  uma amostra aleatória correspondente a  $Y \sim N(\mu, 9)$ .

Assumindo que as duas amostras são independentes, temos que a verossimilhança correspondente à amostra conjunta é dada por

$$\begin{aligned}
 (3.3.2) \quad L(\mu; \mathbf{x}, \mathbf{y}) &= L(\mu; \mathbf{x})L(\mu; \mathbf{y}) \\
 &= \left(\frac{1}{2\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{8}} \left(\frac{1}{3\sqrt{2\pi}}\right)^m e^{-\sum_{i=1}^m \frac{(y_i - \mu)^2}{18}} \\
 &= \left(\frac{1}{2\sqrt{2\pi}}\right)^n \left(\frac{1}{3\sqrt{2\pi}}\right)^m e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{8} - \sum_{i=1}^m \frac{(y_i - \mu)^2}{18}}.
 \end{aligned}$$

Usando o critério da fatoração, não é difícil verificar que uma estatística suficiente para  $\mu$  é dada por

$$(3.3.3) \quad T(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^n X_i}{4} + \frac{\sum_{i=1}^m Y_i}{9}.$$

Além disso, o logaritmo da verossimilhança (3.3.2) pode ser escrito como

$$l(\mu; \mathbf{x}, \mathbf{y}) = -\frac{n}{2} \log 8\pi - \frac{m}{2} \log 18\pi - \sum_{i=1}^n \frac{(x_i - \mu)^2}{8} - \sum_{i=1}^m \frac{(y_i - \mu)^2}{18},$$

de modo que

$$\frac{\partial \log L(\mu; \mathbf{x}, \mathbf{y})}{\partial \mu} = \sum_{i=1}^n \frac{(x_i - \hat{\mu})}{4} + \sum_{i=1}^m \frac{(y_i - \hat{\mu})}{9} = 0,$$

cujas soluções são dadas por

$$\hat{\mu} = \frac{\frac{1}{4} \sum_{i=1}^n X_i + \frac{1}{9} \sum_{i=1}^m Y_i}{\frac{n}{4} + \frac{m}{9}}.$$

Podemos notar que o estimador de máxima verossimilhança é função da estatística suficiente dada em (3.3.3).

### 3.4 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  $\theta = (\theta_1, \dots, \theta_r)$ , ou seja, a verossimilhança depende de dois ou mais parâmetros. O espaço paramétrico será denotado por  $\Theta$ . Nos casos em que as condições de regularidade estão satisfeitas, os estimadores de máxima verossimilhança de  $\theta_1, \dots, \theta_r$  podem ser obtidos como soluções das equações

$$\frac{\partial \log L(\theta; \mathbf{x})}{\partial \theta_i} = 0,$$

$i = 1, \dots, r$ . 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

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

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,

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

conhecida como verossimilhança perfilada de  $\theta_2$  que pode ser usada para que o estimador de máxima verossimilhança de  $\theta_2$  possa ser obtido. A maximização de  $g(\theta_2; \mathbf{x})$  pode, então, ser feita de maneira usual, ou seja, através de derivação, quando possível.

**Exemplo 3.4.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\mu, \sigma^2)$ , onde  $\mu$  e  $\sigma^2$  são desconhecidos. Temos, então, que  $\theta = (\mu, \sigma^2)$ , com

$$L(\theta; \mathbf{x}) = \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}},$$

de modo que

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

Assim

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

que leva ao estimador  $\hat{\mu} = \bar{X}$ . Portanto o logaritmo da verossimilhança perfilada de  $\sigma^2$  é dada por

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

logo 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

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

de modo que os estimadores de máxima verossimilhança de  $\mu$  e  $\sigma^2$  são dados, respectivamente, por

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

No caso multiparamétrico, as mesmas propriedades como invariância, função da estatística suficiente e outras, continuam valendo. O mesmo se aplica ao caso de várias amostras independentes, conforme ilustra o exemplo a seguir.

**Exemplo 3.4.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de  $X \sim N(\mu_x, \sigma^2)$  e  $Y_1, \dots, Y_m$  uma amostra aleatória de  $Y \sim N(\mu_y, \sigma^2)$ . Nesse caso,  $\theta = (\mu_x, \mu_y, \sigma^2)$ . Portanto a verossimilhança correspondente à amostra observada é dada por

$$L(\theta; \mathbf{x}, \mathbf{y}) = \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^n \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^m e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_x)^2 - \frac{1}{2\sigma^2} \sum_{i=1}^m (y_i - \mu_y)^2},$$

logo

$$l(\theta; \mathbf{x}, \mathbf{y}) = -\frac{(n+m)}{2} \log 2\pi - \frac{(m+n)}{2} \log \sigma^2 - \sum_{i=1}^n \frac{(x_i - \mu_x)^2}{2\sigma^2} - \sum_{i=1}^m \frac{(y_i - \mu_y)^2}{2\sigma^2}.$$

Derivando  $l(\theta; \mathbf{x}, \mathbf{y})$  com relação a  $\mu_x$ ,  $\mu_y$  e  $\sigma^2$ , chegamos às equações

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

$$\frac{\partial l(\theta; \mathbf{x}, \mathbf{y})}{\partial \mu_y} = \sum_{j=1}^m (y_j - \hat{\mu}_y) = 0$$

e

$$\frac{\partial l(\theta; \mathbf{x}, \mathbf{y})}{\partial \sigma^2} = -\frac{(m+n)}{2} \frac{1}{\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \left\{ \sum_{i=1}^n (x_i - \hat{\mu}_x)^2 + \sum_{j=1}^m (y_j - \hat{\mu}_y)^2 \right\} = 0,$$

cujas soluções apresentam os estimadores

$$\hat{\mu}_x = \bar{X}, \quad \hat{\mu}_y = \bar{Y}$$

e

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

### 3.5 Família Exponencial e o Método de Máxima Verossimilhança

Se a distribuição da variável aleatória  $X$  pertence à família exponencial unidimensional de distribuições, então o estimador de máxima verossimilhança de  $\theta$  baseado na amostra  $\mathbf{X} = (X_1, \dots, X_n)$  é solução da equação

$$(3.5.1) \quad E[T(\mathbf{X})] = T(\mathbf{X}),$$

desde que a solução pertença ao espaço paramétrico correspondente ao parâmetro  $\theta$ . Esse resultado pode ser estendido para o caso  $k$ -paramétrico em que os estimadores de máxima verossimilhança de  $\theta_1, \dots, \theta_k$  seguem como soluções das equações

$$(3.5.2) \quad E[T_j(\mathbf{X})] = T_j(\mathbf{X}),$$

$j = 1, \dots, k$ .

**Exemplo 3.5.1.** Consideremos uma população com 3 tipos de indivíduos denominados (rotulados) 1, 2, e 3, ocorrendo nas proporções de Hardy-Weinberg

$$p(1; \theta) = \theta^2, \quad p(2; \theta) = 2\theta(1 - \theta), \quad p(3; \theta) = (1 - \theta)^2,$$

$0 < \theta < 1$ . Por exemplo,  $p(1; \theta) = \theta^2$  significa que a probabilidade de se observar um indivíduo do tipo 1 é  $\theta^2$ . Para uma amostra de  $n = 3$  indivíduos, se  $x_1 = 1$ ,  $x_2 = 2$  e  $x_3 = 1$ , onde  $x_1 = 1$  significa que o primeiro indivíduo observado é do tipo 1,  $x_2 = 2$  significa que o segundo indivíduo observado é do tipo 2 e  $x_3 = 1$  significa que o terceiro indivíduo observado é do tipo 1, temos que a função de verossimilhança correspondente é dada por

$$L(\theta; \mathbf{x}) = p(1; \theta)p(2; \theta)p(1; \theta) = 2\theta^5(1 - \theta),$$

de modo que de (3.1.3),

$$l'(\theta; \mathbf{x}) = \frac{5}{\theta} - \frac{1}{1 - \theta} = 0$$

leva ao estimador  $\hat{\theta} = 5/6$  (verifique que  $l''(\hat{\theta}; \mathbf{x}) < 0$ ). Em geral, para uma amostra de  $n$  indivíduos, sendo  $n_1, n_2, n_3$  o número de elementos de  $\{x_1, \dots, x_n\}$  iguais a 1, 2 e 3, respectivamente, temos que

$$L(\theta; \mathbf{x}) = 2^{n_2} \theta^{2n_1+n_2} (1-\theta)^{2n_3+n_2} = 2^{n_2} \left( \frac{\theta}{1-\theta} \right)^{2n_1+n_2} (1-\theta)^{2n}.$$

Então  $c(\theta) = \log(\theta/(1-\theta))$  e  $T(\mathbf{X}) = 2N_1 + N_2$  de modo que

$$E[T(\mathbf{X})] = E[2N_1 + N_2] = 2n\theta^2 + 2n\theta(1-\theta) = 2n\theta.$$

Assim a equação (3.5.1) torna-se

$$2N_1 + N_2 = 2n\hat{\theta}$$

que produz o estimador  $\hat{\theta} = (2N_1 + N_2)/2n$ .

**Exemplo 3.5.2.** Consideremos  $(X_1, Y_1), \dots, (X_n, Y_n)$  uma amostra aleatória da distribuição normal bivariada dada no Exemplo 2.4.4, em que é obtida a estatística suficiente  $\mathbf{T} = (T_1, T_2, T_3, T_4, T_5)$ , com  $T_1 = \sum_{i=1}^n X_i$ ,  $T_2 = \sum_{i=1}^n Y_i$ ,  $T_3 = \sum_{i=1}^n X_i^2$ ,  $T_4 = \sum_{i=1}^n Y_i^2$ ,  $T_5 = \sum_{i=1}^n X_i Y_i$ , para  $\theta = (\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ . Como  $E[X_i] = \mu_x$ ,  $E[Y_i] = \mu_y$ ,  $E[X_i^2] = \mu_x^2 + \sigma_x^2$ ,  $E[Y_i^2] = \mu_y^2 + \sigma_y^2$  e  $E[X_i Y_i] = \mu_x \mu_y + \rho \sigma_x \sigma_y$ ,  $i = 1, \dots, n$ , segue que  $E[T_1] = n\mu_x$ ,  $E[T_2] = n\mu_y$ ,  $E[T_3] = n\mu_x^2 + n\sigma_x^2$ ,  $E[T_4] = n\mu_y^2 + n\sigma_y^2$  e  $E[T_5] = n\mu_x \mu_y + n\rho \sigma_x \sigma_y$ , então de (3.5.2), temos que o estimador de máxima verossimilhança de  $\theta$  tem componentes dadas pelas expressões (2.4.6) e (2.4.7).

### 3.6 O Método dos Momentos

O método dos momentos é um dos métodos de estimação mais simples e antigos. Esse método tem sido utilizado desde o século XVIII. Seja

$$m_r = \frac{1}{n} \sum_{i=1}^n X_i^r,$$

$r \geq 1$ , o  $r$ -ésimo momento amostral de uma amostra aleatória  $X_1, \dots, X_n$ . Seja

$$\mu_r = E[X^r],$$

$r \geq 1$ , o  $r$ -ésimo momento populacional. O método dos momentos consiste na obtenção de estimadores para  $\theta = (\theta_1, \dots, \theta_k)$  resolvendo-se as equações

$$m_r = \mu_r,$$

$r = 1, \dots, k$ .

**Exemplo 3.6.1.** Consideremos novamente o problema da estimação do número de táxis em uma cidade. Sendo  $N$  o número de táxis, vimos que



$$P[X_i = k] = \frac{1}{N}, \quad k = 1, \dots, N,$$

onde  $X_i$  é o número do  $i$ -ésimo táxi observado. Como o primeiro momento populacional é dado por

$$\mu_1 = E[X] = \frac{N+1}{2},$$

temos que um estimador para  $N$ , utilizando-se os primeiros momentos populacional e amostral, é dado pela solução da equação

$$\frac{\hat{N}+1}{2} = \bar{X},$$

de onde segue que

$$\hat{N} = 2\bar{X} - 1.$$

Notemos que, nesse caso, o estimador obtido pelo método dos momentos não é função da estatística suficiente  $X_{(n)}$ .

**Exemplo 3.6.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição da variável aleatória  $X$ , com densidade gama com parâmetros  $\alpha$  e  $\beta$  dados por

$$f(x|\alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \quad x > 0, \alpha > 0, \beta > 0.$$

Como

$$E[X] = \frac{\alpha}{\beta} \quad \text{e} \quad \text{Var}[X] = \frac{\alpha}{\beta^2},$$

temos que estimadores para  $\alpha$  e  $\beta$  são obtidos como solução das equações

$$\frac{\hat{\alpha}}{\hat{\beta}} = \frac{1}{n} \sum_{i=1}^n X_i$$

e

$$\frac{\hat{\alpha}^2}{\hat{\beta}^2} + \frac{\hat{\alpha}}{\hat{\beta}^2} = \frac{1}{n} \sum_{i=1}^n X_i^2$$

que fornece os estimadores

$$(3.6.1) \quad \hat{\alpha} = \frac{\bar{X}^2}{\hat{\sigma}^2}, \quad \text{e} \quad \hat{\beta} = \frac{\bar{X}}{\hat{\sigma}^2},$$

onde  $\hat{\sigma}^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / n$ , como antes. Nesse caso, não é possível obtermos estimadores de máxima verossimilhança explícitos para  $\alpha$  e  $\beta$ . Métodos computacionais como o método do escore considerado na Seção 3.1 devem ser utilizados. Como valores iniciais para esses métodos computacionais, podemos utilizar os estimadores dados por (3.6.1). Notemos também que os estimadores dados por (3.6.1) não são funções da estatística conjuntamente suficiente, que nesse caso é dada por  $(\prod_{i=1}^n X_i, \sum_{i=1}^n X_i)$ .

### 3.7 Estimadores Consistentes

Os métodos de estimação considerados nesta seção produzem, em geral, estimadores consistentes, ou seja, à medida que o tamanho da amostra aumenta, os estimadores ficam tão próximos do parâmetro que está sendo estimado quanto desejado. Consistência está ligada ao conceito de convergência em probabilidade (veja James, 1981).

**Definição 3.7.1.** *Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição da variável aleatória  $X$  que depende do parâmetro  $\theta$ . Dizemos que o estimador  $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$  é consistente para o parâmetro  $\theta$ , se,*

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| > \epsilon) = 0.$$

Em geral, usamos a desigualdade de Chebyshev (veja James, 1981) para a verificação dessa propriedade.

**Exemplo 3.7.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição da variável aleatória  $X$  com média  $\theta$  e variância  $\sigma^2$ . Temos, usando a desigualdade de Chebyshev, que

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

de modo que

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

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

### 3.8 Exercícios

**3.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com função de densidade de probabilidade

$$f(x|\theta) = \frac{\theta}{x^2}, \quad x \geq \theta, \quad \theta > 0.$$

Encontre o estimador de máxima verossimilhança de  $\theta$  e de  $E_\theta[1/X]$ .

**3.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$  com função de densidade de probabilidade dada por

$$f(x|\theta) = \theta x^{\theta-1}, \quad 0 < x < 1, \theta > 0.$$

(i) Encontre os estimadores de máxima verossimilhança de  $\theta$  e de  $g(\theta) = \theta/(1+\theta)$ . (ii) Encontre a distribuição aproximada dos estimadores em (i) quando  $n$  é grande.

**3.3.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\mu, 1)$ . Encontre o estimador de máxima verossimilhança de  $g(\mu) = P_\mu[X > 0]$  e sua distribuição aproximada quando  $n$  é grande.

**3.4.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$  com função de densidade de probabilidade dada por

$$f(x|\theta) = \frac{x}{\theta^2} e^{-x/\theta}, \quad x \geq 0, \theta > 0.$$

- (i) Encontre o estimador de máxima verossimilhança de  $\theta$  e verifique se ele é eficiente.
- (ii) Encontre o estimador de máxima verossimilhança de  $\text{Var}[X]$  e encontre sua distribuição aproximada em grandes amostras.

**3.5.** Encontre a distribuição aproximada para grandes amostras do estimador de máxima verossimilhança de  $\Phi(-\theta)$ , considerado no Exemplo 3.2.2.

**3.6.** Encontre o estimador de máxima verossimilhança de  $\theta^2$  no Exercício 2.9 e compare seu erro quadrático médio com o do estimador eficiente  $\hat{\gamma}$  dado no Exercício 2.9, (i).

**3.7.** Considere uma amostra aleatória de tamanho  $n$  da distribuição da variável aleatória  $X$  onde cada observação apresenta um de três resultados possíveis (por exemplo, favorável, contra e indiferente), que denotamos por “0”, “1” e “2”. Suponhamos que a probabilidade de “0” é  $p_1 = (1 - \theta)/2$ , a probabilidade da ocorrência do resultado “1” é  $p_2 = 1/2$  e do resultado “2” é  $p_3 = \theta/2$ . Seja  $n_1$ : o número de vezes que “0” ocorre,  $n_2$ : o número de vezes que “1” ocorre e  $n_3$ : o número de vezes que o “2” ocorre.

- (i) Encontre, como função de  $n_1, n_2, n_3$ , uma estatística suficiente para  $\theta$ .
- (ii) Encontre o estimador de máxima verossimilhança de  $\theta$ .

**3.8.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$  com função de densidade de probabilidade dada por

$$f(x|\theta) = \theta(\theta + 1)x^{\theta-1}(1 - x), \quad 0 \leq x \leq 1, \theta > 0.$$

- (i) Encontre, usando o método dos momentos, um estimador para  $\theta$ .
- (ii) Encontre o estimador de máxima verossimilhança de  $\theta$  e sua distribuição aproximada em grandes amostras.

**3.9.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável  $X$  com função de densidade de probabilidade dada por

$$f(x|\theta) = \frac{1}{\beta} e^{-\frac{(x-\alpha)}{\beta}} e^{-e^{-\frac{(x-\alpha)}{\beta}}}, \quad -\infty < x < \infty, -\infty < \alpha < \infty, \beta > 0.$$

- (i) Encontre a distribuição de  $Y = e^X$ .
- (ii) Discuta a obtenção do estimador de máxima verossimilhança para  $\beta$ , quando  $\alpha = 0$ .
- (iii) Encontre estatísticas conjuntamente suficientes para  $\alpha$  e  $\beta$ .
- (iv) Discuta a obtenção dos estimadores de máxima verossimilhança para  $\alpha$  e  $\beta$  e verifique se são funções das estatísticas obtidas em (iii).
- (v) Usando (i), gere uma amostra aleatória de tamanho  $n = 20$  da variável aleatória  $Y$ . A partir desta amostra, obtenha uma amostra de tamanho  $n = 20$  para a variável aleatória  $X$  e usando um programa de computador, obtenha os estimadores de máxima verossimilhança de  $\alpha$  e  $\beta$ .

**3.10.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$  com função de densidade de probabilidade

$$f(x|\theta) = \frac{(x+1)}{\theta(\theta+1)} e^{-x/\theta}, \quad x > 0, \theta > 0.$$

- (i) Encontre o estimador de máxima verossimilhança para  $\theta$  e sua distribuição em grandes amostras.
- (ii) Obtenha um estimador para  $\theta$  usando o método dos momentos.

**3.11.** Refaça o Exercício 3.7 supondo agora que  $p_1 = \theta^2$ ,  $p_2 = 2\theta(1 - \theta)$  e  $p_3 = (1 - \theta)^2$ .

**3.12.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição  $N(0, \sigma^2)$ . Encontre o estimador de máxima verossimilhança de  $\sigma$  e sua distribuição em grandes amostras.

**3.13.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com distribuição exponencial com parâmetro  $\theta$ . Encontre o estimador de máxima verossimilhança de  $g(\theta) = P[X > 1]$  e sua distribuição aproximada quando  $n$  for grande.

**3.14.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com função de densidade de probabilidade Weibull dada por

$$f(x|\theta, a) = \theta a x^{a-1} e^{-\theta x^a}; \quad x, a, \theta > 0.$$

- (i) Suponha que  $a$  seja conhecido. Encontre o estimador de máxima verossimilhança de  $\theta$  e sua distribuição aproximada para quando  $n$  for grande.
- (ii) Suponha agora que  $\theta$  e  $a$  são desconhecidos. Encontre as equações de verossimilhança para os dois parâmetros. Proponha um procedimento iterativo para encontrar os estimadores de máxima verossimilhança dos dois parâmetros. Discuta a implementação do procedimento no computador.
- (iii) Gere uma amostra com  $n = 10$  elementos da distribuição de  $X$  assumindo que  $a = \theta = 1$ . Usando o procedimento iterativo em (ii), obtenha estimadores

de máxima verossimilhança de  $a$  e de  $\theta$ . Compare as estimativas com os valores usados para simular a amostra.

**3.15.** Obtenha a informação de Fisher  $I_F(\theta)$  no Exemplo 3.1.6.

**3.16.** Obtenha os estimadores de máxima verossimilhança de  $\beta$  e  $\sigma^2$  no modelo de regressão dado no Exercício 2.12.

**3.17.** Verifique se os estimadores obtidos nos Exemplos 3.1.2, 3.1.3, 3.2.1, 3.2.3 e 3.6.2 são consistentes.

**3.18.** Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias independentes com  $Y_i \sim N(\alpha + \beta x_i, \sigma^2)$ , onde  $x_i$  é conhecido,  $i = 1, \dots, n$ . Encontre os estimadores de máxima verossimilhança de  $\alpha$ ,  $\beta$  e  $\sigma^2$ .

**3.19.** Sejam  $Y_1, \dots, Y_n$  variáveis aleatórias independentes com  $Y_i \sim N(\beta x_i, \sigma^2 x_i)$ , onde  $x_i > 0$  é conhecido,  $i = 1, \dots, n$ . Encontre os estimadores de máxima verossimilhança de  $\beta$  e  $\sigma^2$ .

**3.20.** No caso do modelo do Exercício 3.18, os estimadores de  $\alpha$  e  $\beta$  obtidos através do método de mínimos quadrados minimizam a soma de quadrados  $\sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2$ . Verifique que os estimadores de mínimos quadrados coincidem com os estimadores de máxima verossimilhança de  $\alpha$  e  $\beta$ .

**3.21.** Defina o critério correspondente para obter os estimadores de mínimos quadrados para o modelo do Exercício 3.19.



## 4. Introdução à Teoria das Decisões. Os Princípios Minimax e de Bayes

Neste capítulo apresentamos uma breve introdução à teoria das decisões. Os problemas usuais de estimação e testes de hipóteses são vistos pela ótica da teoria dos jogos, em que os adversários são o estatístico e a natureza. Em primeiro lugar, apresentamos os elementos básicos da teoria das decisões, sendo o objetivo principal a minimização da função de risco. Como, em geral, não é possível a obtenção de um procedimento que minimize a função de risco uniformemente em  $\theta$ , outros critérios para a obtenção de procedimentos ótimos são necessários. Dois desses procedimentos são discutidos neste capítulo. O primeiro é o procedimento minimax, em que o estatístico procura precaver-se contra o risco máximo. A seguir apresentamos o princípio de Bayes em que a característica principal é a formulação do problema de decisão, assumindo que a natureza utiliza um procedimento aleatório, representado por uma distribuição de probabilidade, para escolher um valor para  $\theta$ . Soluções gerais são apresentadas para o estimador de Bayes com respeito a alguns tipos especiais de funções de perda, dentre as quais destacamos a perda quadrática.

### 4.1 Os Elementos Básicos

Os elementos básicos de um problema de decisão são:

- (i) um conjunto não vazio  $\Theta$  dos possíveis estados da natureza que na verdade representa o espaço paramétrico. A natureza escolhe para  $\theta$  um valor nesse conjunto;
- (ii) um conjunto não vazio  $\mathcal{A}$  das possíveis ações que podem ser tomadas pelo estatístico. No caso de problemas de estimação,  $\mathcal{A} = \Theta$ , em geral. No caso de problemas de testes de hipóteses, geralmente  $\mathcal{A}$  consiste nas ações de se aceitar ou rejeitar uma hipótese formulada;
- (iii) uma função  $d : \mathcal{X} \rightarrow \mathcal{A}$ , denominada função de decisão, em que  $\mathcal{X}$  é o espaço amostral associado a uma variável aleatória  $X$  correspondente a um experimento idealizado pelo estatístico para “espionar” (obter informações) sobre a escolha de  $\theta$  feita pela natureza. Seja  $\mathcal{D}$  o conjunto (ou classe) das possíveis funções de decisão. Nessa classe, o estatístico procura um procedimento que seja “melhor”, segundo algum critério;

(iv) uma função real  $l(\theta, a)$ , definida em  $\Theta \times \mathcal{A}$ , que será chamada de função de perda e que satisfaz às seguintes propriedades:

(a)  $l(\theta, a) \geq 0$ , para todo  $\theta \in \Theta$ ,  $a \in \mathcal{A}$ ,

e

(b)  $l(\theta, a) = 0$ , quando  $a = \theta$ ,

ou seja, quando a ação correta é tomada.

Portanto a função  $l(\theta, a)$  representa a perda incorrida pelo estatístico ao tomar a ação  $a$  quando  $\theta$  é a escolha feita pela natureza. Algumas funções de perda comumente empregadas em problemas de decisão são: (i)  $l(\theta, a) = (\theta - a)^2$ , comumente conhecida como perda quadrática; (ii)  $l(\theta, a) = |\theta - a|$ , conhecida como perda do valor absoluto e (iii)  $l(\theta, a) = c(\theta)|\theta - a|^r$ ,  $c(\theta) > 0$ ,  $r > 0$ , que é uma perda mais geral, tendo as perdas em (i) e (ii) como casos particulares.

Como não é possível a implementação de procedimentos que minimizem diretamente a função de perda, pois essa depende de  $\theta$ , que é desconhecido, o estatístico procura minimizar a função de risco, definida a seguir.

**Definição 4.1.1.** *A função de risco correspondente ao procedimento (função de decisão)  $d$  e a função de perda  $l(\theta, a)$  é dada por*

$$(4.1.1) \quad R(\theta, d) = E[l(\theta, d(\mathbf{X}))] = \sum_{\{\mathbf{x} \in \mathcal{X}\}} l(\theta, d(\mathbf{x}))f(\mathbf{x}|\theta),$$

no caso discreto. No caso contínuo, o somatório na expressão acima é substituído por uma integral definida em  $\mathcal{X}$ .

Em (4.1.1),  $f(\mathbf{x}|\theta)$  corresponde à função de verossimilhança da amostra observada (ver Definição 3.1.1). Portanto a função de risco nada mais é do que a perda média sobre o espaço amostral  $\mathcal{X}$ , e é função do parâmetro  $\theta$ . Podemos então comparar procedimentos mediante à utilização da função de risco, conforme definido a seguir.

**Definição 4.1.2.** *Dizemos que um procedimento  $d_1$  é melhor que um procedimento  $d_2$ , quando*

$$(4.1.2) \quad R(\theta, d_1) \leq R(\theta, d_2),$$

para todo  $\theta$ , e

$$(4.1.3) \quad R(\theta, d_1) < R(\theta, d_2),$$

para algum  $\theta$ .

No caso em que (4.1.2) e (4.1.3) estão satisfeitas para todos os procedimentos  $d_2$  em uma certa classe  $\mathcal{D}$  de procedimentos, então dizemos que  $d_1$  é o



melhor procedimento em  $\mathcal{D}$ . Além disso, estando as condições (4.1.2) e (4.1.3) satisfeitas, temos que o procedimento  $d_2$  é dito ser inadmissível. Graficamente, temos a situação da Figura 4.1.

Figura 4.1

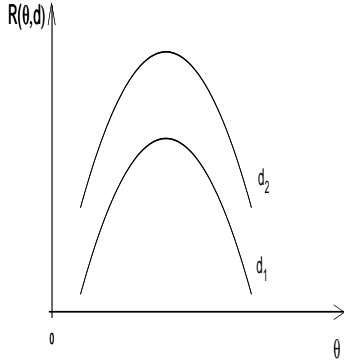
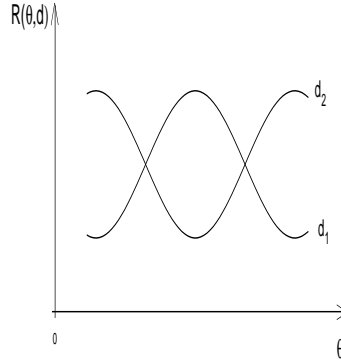


Figura 4.2



Contudo, em geral, ocorre a situação da Figura 4.2, em que o procedimento  $d_1$  é preferível para alguns valores de  $\theta$ , enquanto que para outros valores de  $\theta$ ,  $d_2$  é preferível. Portanto, em geral, não existe um procedimento que seja melhor para todos os valores de  $\theta$ . Em situações como essa, outros critérios devem ser utilizados para se decidir sobre um procedimento em certa classe  $\mathcal{D}$ . O exemplo que apresentamos a seguir ilustra uma tal situação.

**Exemplo 4.1.1.** Suponha que uma moeda apresenta cara com probabilidade  $\theta$  igual a  $1/3$  ou  $2/3$ , ou seja,  $\Theta = \{1/3, 2/3\}$ . É então adequado tomar como espaço das ações  $\mathcal{A} = \{1/3, 2/3\}$ . Para obter informação sobre  $\theta$ , o estatístico faz um lançamento da moeda e observa a variável aleatória  $X$  que denota o número de caras obtidas no lançamento. O espaço amostral associado ao experimento é, portanto,  $\mathcal{X} = \{0, 1\}$ . Nesse caso, podemos definir então quatro funções de decisão,  $d_1, d_2, d_3$  e  $d_4$ , que são dadas por

$$d_1(0) = 1/3, \quad d_2(0) = 1/3, \quad d_3(0) = 2/3, \quad d_4(0) = 2/3,$$

$$d_1(1) = 2/3, \quad d_2(1) = 1/3, \quad d_3(1) = 2/3, \quad d_4(1) = 1/3.$$

Considerando a função de perda do valor absoluto  $l(\theta, a) = |\theta - a|$ , e como a distribuição de  $X$  é discreta, temos que,

$$R(\theta, d) = l(\theta, d(0))P_\theta[X = 0] + l(\theta, d(1))P_\theta[X = 1],$$

onde  $P_\theta[X = 1] = \theta = 1 - P_\theta[X = 0]$ . Portanto, para  $\theta = 1/3$ , temos que

$$\begin{aligned}
R(1/3, d_1) &= l(1/3, d_1(0)).2/3 + l(1/3, d_1(1)).1/3 \\
&= 0.2/3 + 1/3.1/3 = 1/9, \\
R(1/3, d_2) &= 0.2/3 + 0.1/3 = 0, \\
R(1/3, d_3) &= 1/3.2/3 + 1/3.1/3 = 1/3, \\
R(1/3, d_4) &= 1/3.2/3 + 0.1/3 = 2/9.
\end{aligned}$$

Por outro lado, para  $\theta = 2/3$ , de maneira similar, temos que

$$\begin{aligned}
R(2/3, d_1) &= l(2/3, d_1(0)).1/3 + l(2/3, d_1(1)).2/3 \\
&= 1/3.1/3 + 0.2/3 = 1/9, \\
R(2/3, d_2) &= 1/3.1/3 + 1/3.2/3 = 1/3, \\
R(2/3, d_3) &= 0.1/3 + 0.2/9 = 0, \\
R(2/3, d_4) &= 0.1/3 + 1/3.2/3 = 2/9.
\end{aligned}$$

Resumindo os cálculos acima, temos a Tabela 4.1.

**Tabela 4.1.** Riscos de  $d_1, d_2, d_3, d_4$

$d$	$\theta = 1/3$	$\theta = 2/3$	$\max R(\theta; d)$
$d_1$	1/9	1/9	1/9
$d_2$	0	1/3	1/3
$d_3$	1/3	0	1/3
$d_4$	2/9	2/9	2/9

Da Tabela 4.1 podemos concluir que  $R(\theta, d_1) < R(\theta, d_4)$ , para  $\theta = 1/3$  e  $\theta = 2/3$ , de modo que  $d_4$  é inadmissível. Com relação a  $d_1, d_2$  e  $d_3$ , temos a situação da Figura 4.2, em que nenhum procedimento é melhor para todo  $\theta$ .

## 4.2 O Princípio Minimax

Conforme mencionado na introdução, o procedimento minimax é o procedimento que protege o estatístico contra o risco máximo.

**Definição 4.2.1.** Dizemos que o procedimento  $d_0$  é um procedimento minimax numa classe  $\mathcal{D}$  de procedimentos, se

$$\sup_{\theta \in \Theta} R(\theta, d_0) = \inf_{d \in \mathcal{D}} \sup_{\theta \in \Theta} R(\theta, d).$$

Conforme notamos a partir da Definição 4.2.1, o princípio minimax compara simplesmente o máximo dos riscos dos procedimentos.

**Exemplo 4.2.1.** Consideremos novamente a situação do Exemplo 4.1.1. Vimos que o procedimento  $d_4$  é inadmissível. Com relação aos procedimentos  $d_1$ ,  $d_2$  e  $d_3$ , temos da Tabela 4.1 que o procedimento  $d_1$  apresenta o menor risco máximo e, portanto, é o procedimento minimax nesse caso.

**Exemplo 4.2.2.** Seja  $X$  uma única observação de uma variável aleatória  $X$  com distribuição de Poisson com parâmetro  $\theta$ . Portanto consideramos  $\mathcal{A} = \Theta = (0, \infty)$ , com  $\mathcal{X} = \{0, 1, 2, \dots\}$ . Considerando a classe das funções de decisão  $\mathcal{D} = \{d; d(X) = cX\}$ , onde  $c$  é uma constante, temos, para um procedimento  $d$  em  $\mathcal{D}$ , com relação a função de perda

$$l(\theta, a) = \frac{(\theta - a)^2}{\theta},$$

que

$$\begin{aligned} R(\theta, d) &= E[l(\theta, d(X))] \\ &= E\left[\frac{(\theta - cX)^2}{\theta}\right] = \frac{1}{\theta}E[c(X - \theta) + \theta(c - 1)]^2 \\ (4.2.1) \quad &= c^2 + \theta(c - 1)^2. \end{aligned}$$

Como  $R(\theta, d)$  dado em (4.2.1) é uma função linear em  $\theta$  e  $\theta > 0$ , temos que  $R(\theta, d)$  tem máximo finito somente quando  $c = 1$ , pois, nesse caso,  $R(\theta, d) = 1$ , para todo  $\theta$ , ou seja, quando  $c = 1$ ,

$$\max_{\theta \in \Theta} R(\theta, d) = 1.$$

Portanto, na classe  $\mathcal{D}$ ,  $d(X) = X$  é o procedimento minimax.

### 4.3 O Princípio de Bayes

Nesta seção consideramos que a natureza utiliza um mecanismo aleatório para escolher um valor para o parâmetro  $\theta$ . Esse procedimento aleatório é representado por uma distribuição de probabilidade que chamamos de distribuição a priori com função de densidade de probabilidade (ou função de probabilidade, no caso discreto), representada por  $\pi(\theta)$ . Com relação a priori  $\pi$ , temos a seguinte definição.

**Definição 4.3.1.** *O risco de Bayes do procedimento  $d$ , com relação à perda  $l(\theta, d)$  é dado por*

$$r(\pi, d) = E_{\pi}[R(\theta, d)]$$

$$(4.3.1) \quad = \sum_{\{\theta \in \Theta\}} R(\theta, d) \pi(\theta),$$

no caso discreto. No caso em que  $\Theta$  é contínuo, o somatório em (4.3.1) é substituído pela integral correspondente, ou seja,

$$r(\pi, d) = \int_{\Theta} R(\theta, d) \pi(\theta) d\theta.$$

Notemos que se  $R(\theta, d)$  é constante (isto é, independente de  $\theta$ ), então  $r(\pi, d) = R(\theta, d)$ .

**Definição 4.3.2.** Uma função de decisão  $d_B$  é chamada uma função de decisão de Bayes com respeito a priori  $\pi$  e a classe  $\mathcal{D}$  das funções de decisão, se

$$r(\pi, d_B) = \min_{d \in \mathcal{D}} r(\pi, d).$$

**Exemplo 4.3.1.** Consideremos mais uma vez a situação do Exemplo 4.2.1, sendo  $\pi(1/3) = p$  e  $\pi(2/3) = 1 - p$ . De acordo com a Definição 4.3.1, temos que

$$r(\pi, d_1) = \frac{1}{9}\pi(1/3) + \frac{1}{9}\pi(2/3) = \frac{1}{9}p + \frac{1}{9}(1 - p) = 1/9,$$

$$r(\pi, d_2) = 0p + \frac{1}{3}(1 - p) = \frac{1 - p}{3}$$

e

$$r(\pi, d_3) = \frac{1}{3}p + 0(1 - p) = \frac{p}{3}.$$

Portanto temos que, se  $p < 1/3$ ,  $d_3$  é a solução de Bayes. Se  $p = 1/3$ , então  $d_1$  e  $d_3$  são soluções de Bayes. Notemos que nesse caso a solução de Bayes não é única. Se  $1/3 < p < 2/3$ , então  $d_1$  é a solução de Bayes. Se  $p = 2/3$ , então  $d_1$  e  $d_2$  são soluções de Bayes, de modo que nesse caso também a solução de Bayes não é única. Se  $p > 2/3$ , então a solução de Bayes é  $d_2$ .

**Exemplo 4.3.2.** Com relação ao Exemplo 4.2.2, vimos que  $d(X) = X$  é a solução minimax com relação a perda  $l(\theta, a) = (\theta - a)^2/\theta$ . Considerando a priori exponencial com parâmetro um para  $\theta$ , ou seja,

$$\pi(\theta) = e^{-\theta}, \quad \theta > 0,$$

temos que

$$\begin{aligned} r(\pi, d) &= E_{\pi}[R(\theta, d)] = E_{\pi}[c^2 + \theta(c - 1)^2] \\ &= c^2 + (c - 1)^2 E_{\pi}[\theta] \\ &= c^2 + (c - 1)^2. \end{aligned}$$

Como

$$\frac{\partial r(\pi, d)}{\partial c} = 2c + 2(c - 1) = 0,$$

temos que  $r(\pi, d)$  é mínimo quando  $c = 1/2$ , ou seja, com relação a priori e à perda acima, o estimador de Bayes na classe  $\mathcal{D}$  é dado por  $d_B(X) = X/2$ .

#### 4.4 Estimadores de Bayes com Perda Quadrática

Com relação à perda quadrática, é possível a caracterização dos estimadores na classe  $\mathcal{D}$  de todas as funções de decisão. Notemos que no Exemplo 4.3.2, o estimador de Bayes foi obtido numa particular classe de estimadores, ou seja,  $\mathcal{D} = \{d; d(X) = cX\}$ . Contudo a função de perda não era quadrática. O resultado para perda quadrática é enunciado e provado a seguir para o caso em que  $X$  é uma variável aleatória contínua.

**Teorema 4.4.1.** *Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição da variável aleatória  $X$ , com função de densidade de probabilidade  $f(x|\theta)$ . Consideremos para  $\theta$  a distribuição a priori com função de densidade de probabilidade  $\pi(\theta)$ . Então, com relação à perda quadrática, o procedimento (estimador) de Bayes na classe  $\mathcal{D}$  de todas as funções de decisão é dado por*

$$d_B(\mathbf{X}) = E[\theta|\mathbf{X}],$$

ou seja, é o valor esperado de  $\theta$  calculado na distribuição condicional de  $\theta$  dado  $X_1, \dots, X_n$ , que é denominada “distribuição a posteriori de  $\theta$ ”.

**Prova.** Com relação à perda quadrática, a função de risco de um procedimento qualquer  $d(\mathbf{X})$  é dada por

$$(4.4.1) \quad R(\theta, d) = \int_{\mathcal{X}} (\theta - d(\mathbf{x}))^2 f(\mathbf{x}|\theta) d\mathbf{x},$$

onde  $\mathbf{x} = (x_1, \dots, x_n)$ ,  $\mathcal{X}$  é o espaço amostral e  $f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta)$  é a função de verossimilhança correspondente à amostra observada. Com relação a priori  $\pi$ , temos de (4.4.1) que o risco de Bayes do procedimento  $d(\mathbf{X})$  é dado por

$$(4.4.2) \quad \begin{aligned} r(\pi, d) &= \int_{\Theta} \left[ \int_{\mathcal{X}} (d(\mathbf{x}) - \theta)^2 f(\mathbf{x}|\theta) d\mathbf{x} \right] \pi(\theta) d\theta \\ &= \int_{\Theta} \int_{\mathcal{X}} (d(\mathbf{x}) - \theta)^2 f(\mathbf{x}|\theta) \pi(\theta) d\mathbf{x} d\theta. \end{aligned}$$

Como

$$(4.4.3) \quad f(\mathbf{x}|\theta)\pi(\theta) = f(\mathbf{x};\theta) = \pi(\theta|\mathbf{x})g(\mathbf{x}),$$

temos de (4.4.2) que

$$(4.4.4) \quad \begin{aligned} r(\pi, d) &= \int_{\Theta} \int_{\mathcal{X}} (d(\mathbf{x}) - \theta)^2 \pi(\theta|\mathbf{x}) g(\mathbf{x}) d\mathbf{x} d\theta \\ &= \int_{\mathcal{X}} \left[ \int_{\Theta} (d(\mathbf{x}) - \theta)^2 \pi(\theta|\mathbf{x}) d\theta \right] g(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

De acordo com a Definição 4.3.2, temos que o procedimento de Bayes é o procedimento que minimiza (4.4.4), ou seja, para cada  $\mathbf{x}$ , é o procedimento que minimiza

$$(4.4.5) \quad \int_{\Theta} (d(\mathbf{x}) - \theta)^2 \pi(\theta|\mathbf{x}) d\theta = E[(d(\mathbf{X}) - \theta)^2 | \mathbf{X}].$$

Derivando (4.4.5) com relação a  $d(\mathbf{X})$  e igualando a derivada a zero, chegamos ao procedimento

$$d_B(\mathbf{X}) = E[\theta | \mathbf{X}],$$

que é a forma geral do estimador de Bayes com relação à perda quadrática.

De (4.4.3) temos que

$$(4.4.6) \quad \pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)}{g(\mathbf{x})} = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{g(\mathbf{x})},$$

onde

$$(4.4.7) \quad g(\mathbf{x}) = \int_{\Theta} f(\mathbf{x}|\theta)\pi(\theta) d\theta$$

é a densidade marginal de  $\mathbf{x} = (x_1, \dots, x_n)$ . A densidade  $\pi(\theta|\mathbf{x})$  é denominada função de densidade de probabilidade a posteriori e pode ser interpretada diretamente a partir do Teorema de Bayes, ou seja, a densidade (ou função de probabilidade) condicional é igual à densidade (ou função de probabilidade) conjunta dividida pela densidade (ou função de probabilidade) marginal de  $\mathbf{x}$ . O Teorema 4.4.1 pode ser generalizado para o caso de uma função qualquer de  $\theta$ ,  $\tau(\theta)$ , ou seja, o estimador de Bayes de  $\tau(\theta)$  com relação à perda quadrática é dado por

$$d_B(\mathbf{x}) = E[\tau(\theta) | \mathbf{X}] = \int_{\Theta} \tau(\theta) \pi(\theta|\mathbf{x}) d\theta.$$

Notemos, portanto, que os estimadores de Bayes não são invariantes, como são os estimadores de máxima verossimilhança no sentido de que sendo  $\hat{\theta}$  um

estimador de Bayes de  $\theta$ ,  $\tau(\hat{\theta})$  não é necessariamente um estimador de Bayes de  $\tau(\theta)$ .

**Exemplo 4.4.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$  com distribuição de Bernoulli com parâmetro  $\theta$ . Consideremos para  $\theta$  a função de densidade a priori

$$\pi(\theta) = \frac{\Gamma[a+b]}{\Gamma[a]\Gamma[b]} \theta^{a-1} (1-\theta)^{b-1},$$

$0 < \theta < 1$ ,  $a, b > 0$ , usualmente conhecida como densidade beta com parâmetros  $a$  e  $b$ , que denotamos por  $Beta(a, b)$  e onde  $\Gamma[a]$  é a função gama avaliada no ponto  $a$ , ou seja,

$$(4.4.8) \quad \Gamma[a] = \int_0^\infty x^{a-1} e^{-x} dx.$$

Como

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta) = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i},$$

temos de (4.4.7) que,

$$\begin{aligned} g(\mathbf{x}) &= \int_0^1 \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i} \frac{\Gamma[a+b]}{\Gamma[a]\Gamma[b]} \theta^{a-1} (1-\theta)^{b-1} d\theta \\ &= \frac{\Gamma[a+b]}{\Gamma[a]\Gamma[b]} \int_0^1 \theta^{\sum_{i=1}^n x_i + a - 1} (1-\theta)^{n-\sum_{i=1}^n x_i + b - 1} d\theta \\ &= \frac{\Gamma[a+b]}{\Gamma[a]\Gamma[b]} \frac{\Gamma[\sum_{i=1}^n x_i + a] \Gamma[n - \sum_{i=1}^n x_i + b]}{\Gamma[n + a + b]}. \end{aligned}$$

Portanto de (4.4.6) temos que

$$\begin{aligned} \pi(\theta|\mathbf{x}) &= \frac{\frac{\Gamma[a+b]}{\Gamma[a]\Gamma[b]} \theta^{\sum_{i=1}^n x_i + a - 1} (1-\theta)^{n-\sum_{i=1}^n x_i + b - 1}}{\frac{\Gamma[a+b]}{\Gamma[a]\Gamma[b]} \frac{\Gamma[\sum_{i=1}^n x_i + a] \Gamma[n - \sum_{i=1}^n x_i + b]}{\Gamma[n + a + b]}} \\ &= \frac{\Gamma[n + a + b]}{\Gamma[\sum_{i=1}^n x_i + a] \Gamma[n - \sum_{i=1}^n x_i + b]} \theta^{\sum_{i=1}^n x_i + a - 1} (1-\theta)^{n-\sum_{i=1}^n x_i + b - 1}, \end{aligned}$$

ou seja, a distribuição a posteriori de  $\theta$  dado  $\mathbf{X}$  é uma distribuição beta com parâmetros  $\sum_{i=1}^n x_i + a$  e  $n - \sum_{i=1}^n x_i + b$  que denotamos por

$$\theta|\mathbf{X} \sim Beta\left(\sum_{i=1}^n x_i + a; n - \sum_{i=1}^n x_i + b\right).$$

Então, o estimador de Bayes de  $\theta$  com relação à perda quadrática é dado por

$$(4.4.9) \quad d_B(\mathbf{X}) = E[\theta|\mathbf{X}] = \frac{\sum_{i=1}^n x_i + a}{n + a + b}.$$

Notemos, dos cálculos acima, que as distribuições a priori e a posteriori pertencem à mesma família de distribuições, ou seja, no caso em que a distribuição de  $X$  é Bernoulli e a distribuição a priori é da família Beta, a distribuição a posteriori é também da família Beta. Dizemos, então, que a distribuição Beta é conjugada para a Bernoulli. É também verdade que a distribuição Beta é conjugada para as distribuições Binomial e Binomial Negativa. Os parâmetros  $a$  e  $b$  da priori beta devem ser escolhidos de modo que  $\pi(\theta)$  expresse o conhecimento a priori que o estatístico tem sobre  $\theta$ . No caso particular em que  $a = b = 1$ , temos que

$$(4.4.10) \quad \pi(\theta) = 1, \quad 0 < \theta < 1,$$

ou seja, nesse caso a distribuição  $U(0, 1)$  é escolhida como priori para  $\theta$ . No caso da priori uniforme, temos de (4.4.9) que

$$(4.4.11) \quad d_B(\mathbf{X}) = \frac{\sum_{i=1}^n X_i + 1}{n + 2}.$$

A priori uniforme indica que, inicialmente, o estatístico tem pouca informação sobre  $\theta$ , pois com relação a essa priori, qualquer intervalo de mesmo comprimento tem a mesma área (probabilidade).

Para calcularmos o risco de Bayes do estimador (4.4.11) com relação a priori uniforme, temos que

$$\begin{aligned} R(\theta, d) &= E \left[ \left( \frac{\sum_{i=1}^n X_i + 1}{n + 2} - \theta \right)^2 \right] \\ &= \frac{1}{(n + 2)^2} E \left[ \left( \sum_{i=1}^n X_i - n\theta + 1 - 2\theta \right)^2 \right] \\ &= \frac{1}{(n + 2)^2} [(4 - n)\theta^2 - (4 - n)\theta + 1]. \end{aligned}$$

Com relação a priori uniforme dada em (4.4.10), temos que  $E_\pi[\theta] = 1/2$ ,  $Var_\pi[\theta] = 1/12$  e  $E_\pi[\theta^2] = 1/3$ , de modo que

$$r(\pi, d) = \frac{1}{(n + 2)^2} \left[ \frac{(4 - n)}{3} - \frac{(4 - n)}{2} + 1 \right]$$



$$= \frac{1}{6(n+2)}.$$

Certamente, o estimador de Bayes em (4.4.11) tem risco de Bayes menor, com relação a priori uniforme acima, que o risco de Bayes do estimador de máxima verossimilhança  $\hat{\theta} = \bar{X}$ .

**Exemplo 4.4.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição da variável aleatória  $X$  com distribuição de  $Poisson(\theta)$ . Consideremos para  $\theta$  a distribuição a priori com função de densidade de probabilidade

$$(4.4.12) \quad \pi(\theta) = \frac{b^a \theta^{a-1} e^{-\theta b}}{\Gamma[a]},$$

$\theta > 0$ ,  $a > 0$ ,  $b > 0$ , ou seja, gama com parâmetros  $a$  e  $b$ , que denotamos por  $Gama(a, b)$ . Em (4.4.12),  $\Gamma[a]$  é como definido em (4.4.8). Como

$$\begin{aligned} f(\mathbf{x}|\theta)\pi(\theta) &= \frac{\frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \theta^{a-1} e^{-\theta b} b^a}{\Gamma[a]} \\ &= \frac{b^a e^{-\theta(n+b)} \theta^{\sum_{i=1}^n x_i + a - 1}}{\prod_{i=1}^n x_i! \Gamma[a]}, \end{aligned}$$

$\theta > 0$ , temos que

$$\begin{aligned} g(\mathbf{x}) &= \int_0^\infty \frac{b^a e^{-\theta(n+b)} \theta^{\sum_{i=1}^n x_i + a - 1}}{\prod_{i=1}^n x_i! \Gamma[a]} d\theta \\ &= \frac{b^a}{\prod_{i=1}^n x_i! \Gamma[a]} \frac{\Gamma[\sum_{i=1}^n x_i + a]}{(n+b) \sum_{i=1}^n x_i + a}. \end{aligned}$$

Portanto

$$\pi(\theta|\mathbf{x}) = \frac{e^{-\theta(n+b)} \theta^{\sum_{i=1}^n x_i + a - 1}}{\frac{\Gamma[\sum_{i=1}^n x_i + a]}{(n+b) \sum_{i=1}^n x_i + a}},$$

ou seja, a distribuição a posteriori de  $\theta$  dado  $\mathbf{X}$  é uma distribuição gama com parâmetros  $\sum_{i=1}^n x_i + a$  e  $n + b$  que denotamos por

$$\theta|\mathbf{X} \sim \Gamma \left[ \sum_{i=1}^n x_i + a; n + b \right].$$

Assim,

$$E[\theta|\mathbf{X}] = \frac{\sum_{i=1}^n x_i + a}{n + b}.$$

Além disso, no caso da Poisson, como visto acima, priori gama leva a uma posteriori gama, de modo que a distribuição gama é conjugada para a Poisson. Após algumas manipulações algébricas, não é difícil verificar que (ver Exercício 4.5)

$$\begin{aligned} R(\theta, d) &= E \left[ \left( \frac{\sum_{i=1}^n x_i + a}{n + b} - \theta \right)^2 \right] \\ &= \frac{1}{(n + b)^2} [a^2 + b^2 \theta^2 + \theta(n - 2ab)], \end{aligned}$$

de modo que

$$r(\pi, d) = E_\pi[R(\theta, d)] = \frac{a}{b(n + b)}.$$

**Exemplo 4.4.3.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$  com distribuição  $N(\mu, \sigma_0^2)$ , onde  $\sigma_0^2$  é conhecido. Consideremos para  $\mu$  a priori  $N(a, b^2)$ , ou seja,

$$\pi(\mu) = \frac{1}{\sqrt{2\pi}b} e^{-\frac{(\mu-a)^2}{2b^2}},$$

onde  $a$  e  $b$  são conhecidos. A priori  $N(a, b^2)$  expressa o fato de que  $a$  é um valor razoável para  $\mu$  enquanto que  $b^2$  (ou  $b$ ) quantifica a confiança (ou certeza) de que  $a$  é um valor razoável para  $\mu$ . Quanto maior  $b^2$  (ou  $b$ ), mais incerto o estatístico está com relação a escolha feita pela natureza com relação a  $\mu$ . Após uma série de manipulações algébricas (verifique!), temos que

$$\begin{aligned} f(\mathbf{x}|\mu)\pi(\mu) &= \left( \frac{1}{\sqrt{2\pi}\sigma_0} \right)^n \frac{1}{\sqrt{2\pi}b} e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma_0^2} - \frac{(\mu - a)^2}{2b^2}} \\ &= \left( \frac{1}{\sqrt{2\pi}\sigma_0} \right)^n \frac{1}{\sqrt{2\pi}b} e^{-\sum_{i=1}^n \frac{x_i^2}{2\sigma_0^2} - \frac{a^2}{2b^2} + \frac{\left( \sum_{i=1}^n \frac{x_i}{\sigma_0^2} + \frac{a}{b^2} \right)^2}{2 \left( \frac{n}{\sigma_0^2} + \frac{1}{b^2} \right)}} \times \\ &\quad \times e^{-\frac{1}{2} \left( \frac{n}{\sigma_0^2} + \frac{1}{b^2} \right) \left( \mu - \frac{\sum_{i=1}^n \frac{x_i}{\sigma_0^2} + \frac{a}{b^2}}{\frac{n}{\sigma_0^2} + \frac{1}{b^2}} \right)^2}, \end{aligned}$$

e

$$g(\mathbf{x}) = \left( \frac{1}{\sqrt{2\pi}\sigma_0} \right)^n \frac{1}{b} \sqrt{\frac{1}{\frac{n}{\sigma_0^2} + \frac{1}{b^2}}} e^{-\sum_{i=1}^n \frac{x_i^2}{2\sigma_0^2} - \frac{a^2}{2b^2} + \frac{\left( \sum_{i=1}^n \frac{x_i}{\sigma_0^2} + \frac{a}{b^2} \right)^2}{2 \left( \frac{n}{\sigma_0^2} + \frac{1}{b^2} \right)}},$$

de modo que (verifique!)

$$\pi(\mu|\mathbf{x}) = \frac{1}{\sqrt{2\pi} \sqrt{\frac{1}{\frac{n}{\sigma_0^2} + \frac{1}{b^2}}}} e^{-\frac{1}{2} \left( \frac{n}{\sigma_0^2} + \frac{1}{b^2} \right) \left( \mu - \frac{\frac{\sum_{i=1}^n x_i}{\sigma_0^2} + \frac{a}{b^2}}{\frac{n}{\sigma_0^2} + \frac{1}{b^2}} \right)^2},$$

ou seja, a distribuição a posteriori de  $\mu$  dado  $X_1, \dots, X_n$  é normal com média  $(\sum_{i=1}^n x_i/\sigma_0^2 + a/b^2)/(\frac{n}{\sigma_0^2} + 1/b^2)$  e variância  $1/(\frac{n}{\sigma_0^2} + 1/b^2)$ , que denotamos por

$$\mu|\mathbf{X} \sim N \left( \frac{\sum_{i=1}^n \frac{x_i}{\sigma_0^2} + \frac{a}{b^2}}{\frac{n}{\sigma_0^2} + \frac{1}{b^2}}; \frac{1}{\frac{n}{\sigma_0^2} + \frac{1}{b^2}} \right).$$

Temos, portanto, que a priori normal é conjugada para a distribuição normal quando queremos estimar  $\mu$  com variância  $\sigma_0^2$  conhecida. Com relação a perda quadrática, temos, então, que

$$d_B = \frac{\frac{\sum_{i=1}^n X_i}{\sigma_0^2} + \frac{a}{b^2}}{\frac{n}{\sigma_0^2} + \frac{1}{b^2}} = \frac{\frac{n}{\sigma_0^2}}{\frac{n}{\sigma_0^2} + \frac{1}{b^2}} \bar{X} + \frac{\frac{1}{b^2}}{\frac{n}{\sigma_0^2} + \frac{1}{b^2}} a,$$

de modo que o estimador de Bayes de  $\mu$  é uma combinação linear convexa (coeficientes somam um) entre a média amostral  $\bar{X}$  (que é o estimador eficiente e de máxima verossimilhança de  $\mu$ ) e a média a priori  $a$ . Notemos que quanto maior  $n$ , maior o peso atribuído a  $\bar{X}$ . Portanto para  $n$  grande a distribuição a priori tem pouca influência na distribuição a posteriori. Por outro lado, valores pequenos de  $b$  aumentam a contribuição de  $a$  no estimador  $d_B$  acima. Lembramos que  $b$  pequeno indica uma maior confiança do estatístico de que  $a$  é um valor razoável para  $\mu$ . Temos também que (verifique!)

$$R(\mu, d_B) = E \left[ \left( \frac{\frac{n\bar{X}}{\sigma_0^2} + \frac{a}{b^2}}{\frac{n}{\sigma_0^2} + \frac{1}{b^2}} - \mu \right)^2 \right] = \frac{\frac{n}{\sigma_0^2} + \frac{(a-\mu)^2}{b^4}}{\left( \frac{n}{\sigma_0^2} + \frac{1}{b^2} \right)^2}$$

e

$$r(\pi, d) = \frac{1}{\frac{n}{\sigma_0^2} + \frac{1}{b^2}}.$$

Além disso,

$$R(\mu, d_B) \rightarrow \frac{\sigma_0^2}{n},$$

quando  $b \rightarrow \infty$ , ou seja, a informação a priori é pouco precisa.

Para finalizar o capítulo, apresentamos a seguir um resultado importante, relacionando os estimadores de Bayes a uma estatística suficiente.

**Teorema 4.4.2.** *Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição da variável aleatória  $X$  com função de densidade (ou de probabilidade)  $f(x|\theta)$ . Seja  $T = T(X_1, \dots, X_n)$  uma estatística suficiente para  $\theta$ . Consideremos para  $\theta$  a função de densidade (ou de probabilidade)  $\pi(\theta)$ . Então, o estimador de Bayes de  $\theta$  com relação à perda quadrática é função de  $T$ .*

**Prova.** Vamos considerar a demonstração apenas para o caso em que  $X$  e  $\theta$  são variáveis aleatórias contínuas. Sendo  $T$  uma estatística suficiente para  $\theta$ , usando o Critério da Fatoração, podemos escrever

$$f(\mathbf{x}|\theta) = h(\mathbf{x})g_\theta(t(\mathbf{x})),$$

ou seja,  $g_\theta(t(\mathbf{x}))$  depende de  $\mathbf{x}$  somente por  $t(\mathbf{x})$ . Podemos, então, escrever a função de densidade (ou de probabilidade) a posteriori como

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{\Theta} f(\mathbf{x}|\theta)\pi(\theta)d\theta}$$

$$\frac{h(\mathbf{x})g_\theta(t(\mathbf{x}))\pi(\theta)}{\int_{\Theta} h(\mathbf{x})g_\theta(t(\mathbf{x}))\pi(\theta)d\theta} = \frac{g_\theta(t(\mathbf{x}))\pi(\theta)}{\int_{\Theta} g_\theta(t(\mathbf{x}))\pi(\theta)d\theta},$$

de modo que a função de densidade a posteriori depende de  $\mathbf{x}$  somente através de  $T = T(\mathbf{x})$ . Como o estimador de Bayes de  $\theta$  com relação à perda quadrática é a média da posteriori, ele dependerá de  $\mathbf{X}$  somente através de  $T$ .

O resultado do Teorema 4.4.2 vale na verdade em situações mais gerais no que diz respeito à função de perda. Na verdade qualquer que seja a função de perda considerada, o estimador de Bayes só dependerá de  $\mathbf{X}$  através de  $T = T(X_1, \dots, X_n)$ , pois qualquer que seja a função de perda, o estimador de Bayes é obtido utilizando a distribuição a posteriori  $\pi(\theta|\mathbf{x})$ .

## 4.5 Exercícios

**4.1.** Seja  $X$  uma única observação da distribuição  $N(\mu, 1)$ , onde  $-\infty < \mu < \infty$ . Considere a perda quadrática.

- (i) Encontre o risco  $R(\mu, d)$  para a classe  $\mathcal{D} = \{d; d(x) = cX\}$ .
- (ii) Encontre, na classe  $\mathcal{D}$ , o estimador minimax de  $\mu$ .
- (iii) Encontre em  $\mathcal{D}$  o estimador de Bayes de  $\mu$  com relação a priori  $\pi(\mu) = 1/2; -1 \leq \mu \leq 1$ .

**4.2.** Seja  $X$  uma única observação da variável aleatória  $X$  com função de probabilidade

$$f(x|\theta) = \frac{2!}{x!(2-x)!} \theta^x (1-\theta)^{2-x}, \quad x = 0, 1, 2,$$

onde  $0 < \theta < 1$ . Considere os estimadores  $d_1(X) = X/2$  e  $d_2(X) = (X+1)/4$  e função de perda quadrática.

- (i) Verifique se existe um estimador uniformemente melhor (melhor para todo  $\theta$ ), ou seja, verifique se um dos estimadores é inadmissível.
- (ii) Qual dos estimadores é minimax?

**4.3.** Considere uma única observação da variável aleatória  $X \sim \text{Binomial}(m, \theta)$ . Seja  $l(\theta, d) = (\theta - d)^2$ .

- (i) Encontre o risco de  $d(X) = X/m$ .
- (ii) Encontre o risco de Bayes de  $d(X)$  em (i), com relação a priori  $\pi(\theta) = 1, 0 \leq \theta \leq 1$ .

**4.4.** Refaça o Exercício 4.3., considerando agora a perda  $l(\theta, d) = (\theta - d)^2 / \theta(1 - \theta)$ .

**4.5.** Seja uma única observação da distribuição  $\text{Poisson}(\theta)$ . Encontre o risco de Bayes do estimador  $d(X) = X$ , com relação à perda quadrática e a priori  $\text{Gamma}(\alpha, \beta)$ .

**4.6.** Considere o problema de se estimar  $\theta \in \Theta = \{0, 1\}$ , baseado em uma única observação da variável aleatória  $X$ , com densidade

$$f(x|\theta) = 2^{-(x+\theta)}, \quad x = 1 - \theta, 2 - \theta, 3 - \theta, \dots$$

Considere a perda 0-1, ou seja,

$$l(0, 0) = l(1, 1) = 0 \quad \text{e} \quad l(0, 1) = l(1, 0) = 1.$$

Considere também os estimadores

$$d_1(X) = \begin{cases} 1, & X = 0, \\ 0, & X > 0, \end{cases} \quad \text{e} \quad d_2(X) = \begin{cases} 0, & X \leq 1, \\ 1, & X > 1, \end{cases}$$

- (i) Encontre  $R(\theta, d_i(X)), i = 1, 2$ .
- (ii) Qual dos estimadores é minimax? Alguns dos estimadores é inadmissível?

**4.7.** Seja  $X$  uma única observação da distribuição  $U(0, \theta)$ , onde  $\theta$  é uma variável aleatória com densidade

$$\pi(\theta) = \theta e^{-\theta}, \quad \theta > 0.$$

- (i) Encontre a densidade a posteriori de  $\theta$ .
- (ii) Encontre o estimador de Bayes de  $\theta$  com respeito à perda quadrática.

**4.8.** Seja  $X$  o tempo de vida de uma lâmpada (em mil horas) fabricada por certa companhia. Considera-se que  $X$  é uma variável aleatória com densidade

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

Considere para  $\theta$  a priori

$$\pi(\theta) = 16\theta e^{-4\theta}, \quad \theta > 0.$$

- (i) Encontre a distribuição a posteriori de  $\theta$ .
- (ii) Encontre o estimador de Bayes de  $E(X)$  e  $Var(X)$  com relação à perda quadrática.

**4.9.** Em uma área de reflorestamento, o número de árvores de determinada espécie, por hectare, com certa doença tem uma distribuição  $Poisson(\theta)$ . A distribuição a priori de  $\theta$  é exponencial com média igual a 1. Encontre o estimador de Bayes de  $P_\theta(X = 0)$  com relação à perda quadrática..

**4.10.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição  $U(0, \theta)$ . Suponhamos que  $\theta$  seja uma variável aleatória com função de densidade de probabilidade (Pareto)

$$\pi(\theta) = \begin{cases} ba^b/\theta^{b+1}, & \theta \geq a, \\ 0, & \theta < a, \end{cases}$$

Encontre a distribuição a posteriori de  $\theta$  e o estimador de Bayes de  $\theta$  com relação à perda quadrática.

**4.11.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim Bernoulli(\theta)$ . Considere para  $\theta$  a priori

$$\pi(\theta) = \begin{cases} 2\theta, & 0 < \theta < 1, \\ 0, & \text{caso contrário,} \end{cases}$$

Encontre o estimador de Bayes de  $\theta$  com relação à perda quadrática e seu risco de Bayes.

**4.12.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da densidade

$$f(x|\theta) = \theta x^{\theta-1}, \quad 0 < x < 1, \quad \theta > 0.$$

Vamos assumir para  $\theta$  a priori gama

$$\pi(\theta) = \lambda^r \theta^{r-1} e^{-\theta\lambda} / \Gamma(r),$$

onde  $r$  e  $\lambda$  são conhecidos. Encontre a distribuição a posteriori de  $\theta$  e o estimador de Bayes de  $\theta$  com relação à perda quadrática.

## 5. Estimação por Intervalo

Neste capítulo consideramos o problema de estimação de parâmetros utilizando intervalos de confiança. Os intervalos clássicos são obtidos a partir de variáveis aleatórias especiais que denominamos quantidades pivotais. Os intervalos de confiança Bayesianos são obtidos utilizando a distribuição a posteriori. Em primeiro lugar, discutimos propriedades da média e da variância amostrais quando as amostras são obtidas a partir de populações normais. A seguir introduzimos os métodos de construção de intervalos.

### 5.1 Amostras de Populações Normais

Os resultados que apresentamos a seguir são utilizados com bastante frequência na construção de intervalos de confiança e testes de hipóteses para populações normais.

**Teorema 5.1.** *Sejam  $X_1, \dots, X_n$  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)  $\frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1}$ ;*

*onde  $\chi_\nu^2$  denota uma variável aleatória com distribuição quiquadrado com  $\nu$  graus de liberdade, isto é, com f.d.p. dada por*

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

*$t_\nu$  denota uma variável aleatória com distribuição  $t$  de Student com  $\nu$  graus de liberdade, isto é, com f.d.p. dada por*

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

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

**Prova.** (i) Temos que

$$\bar{X} \sim N(\mu, \sigma^2/n),$$

enquanto que  $X_i - \bar{X} \sim N\left(0, \sigma^2 \frac{(n-1)}{n}\right)$ . Por outro lado, a função geradora de momentos (James, 1981) de  $Y_1 = \bar{X}$  e  $Y_2 = X_i - \bar{X}$  é dada por

$$\begin{aligned} M_{Y_1, Y_2}(s_1, s_2) &= E \left[ e^{s_1 \bar{X} + s_2 (X_i - \bar{X})} \right] = E \left[ e^{s_2 X_i + \bar{X}(s_1 - s_2)} \right] \\ &= E \left[ e^{(s_2 + \frac{(s_1 - s_2)}{n}) X_i + \frac{(s_1 - s_2)}{n} \sum_{j \neq i}^n X_j} \right] \\ &= E \left[ e^{(s_2 + \frac{(s_1 - s_2)}{n}) X_i} \right] E \left[ e^{\frac{(s_1 - s_2)}{n} \sum_{j \neq i}^n X_j} \right]. \end{aligned}$$

Como  $X_i \sim N(\mu, \sigma^2)$  e  $\sum_{j \neq i}^n X_j \sim N((n-1)\mu, (n-1)\sigma^2)$ , temos que

$$\begin{aligned} M_{Y_1, Y_2}(s_1, s_2) &= e^{\mu \left( s_2 + \frac{(s_1 - s_2)}{n} \right) + \frac{\sigma^2}{2} \left( s_2 + \frac{(s_1 - s_2)}{n} \right)^2} \\ &\quad \times e^{\frac{(n-1)}{n} (s_1 - s_2) \mu + \frac{1}{2} \left\{ \left( \frac{s_1 - s_2}{n} \right)^2 (n-1) \sigma^2 \right\}} \\ &= e^{\mu s_1 + \frac{s_1^2 \sigma^2}{2n}} e^{\frac{s_2^2 (n-1) \sigma^2}{2n}} \end{aligned}$$

que é o produto das funções geradoras de momentos das distribuições de  $\bar{X}$  e  $X_i - \bar{X}$ . Portanto temos que  $X_i - \bar{X}$  e  $\bar{X}$  são independentes, pois a função geradora da distribuição conjunta é o produto das funções geradoras de momentos das distribuições marginais. Como  $\sum_{i=1}^n (X_i - \bar{X})^2$  é função de  $X_i - \bar{X}$  que é independente de  $\bar{X}$ , temos que  $S^2$  é independente de  $\bar{X}$ .

(ii) Não é difícil verificar que

$$(5.1.1) \quad \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\sigma^2} + n \frac{(\bar{X} - \mu)^2}{\sigma^2}.$$

Como  $(X_i - \mu)/\sigma \sim N(0, 1)$ , temos que  $(X_i - \mu)^2/\sigma^2 \sim \chi_1^2$ ,  $i = 1, \dots, n$ , de modo que

$$Y_1 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi_n^2.$$

Também  $n(\bar{X} - \mu)^2/\sigma^2 \sim \chi_1^2$ . Como a função geradora de momentos da distribuição quiquadrado com  $g$  graus de liberdade é dada por

$$M_g(s) = (1 - 2s)^{-g/2},$$



temos que as funções geradoras das distribuições quiquadrado com  $g = 1$  e  $g = n$  graus de liberdade são dadas respectivamente por

$$(5.1.2) \quad M_1(s) = (1 - 2s)^{-1/2} \quad \text{e} \quad M_n(s) = (1 - 2s)^{-n/2}.$$

Além disso, como  $\bar{X}$  e  $S^2$  são independentes, temos que os dois termos do lado direito de (5.1.1) que denotamos por  $Y_2$  e  $Y_3$ , respectivamente, são independentes, de modo que

$$M_{Y_1}(s) = M_{Y_2}(s)M_{Y_3}(s),$$

ou seja, de (5.1.2) segue que

$$M_{Y_2}(s) = \frac{M_{Y_1}(s)}{M_{Y_3}(s)} = (1 - 2s)^{-(n-1)/2},$$

logo a distribuição de  $Y_2 = (n-1)S^2/\sigma^2$  é quiquadrado com  $n-1$  graus de liberdade.

(iii) Note que podemos escrever

$$(5.1.3) \quad \sqrt{n} \frac{(\bar{X} - \mu)}{S} = \frac{\sqrt{n} \frac{(\bar{X} - \mu)}{\sigma}}{\sqrt{\frac{(n-1)S^2}{(n-1)\sigma^2}}}$$

que corresponde ao quociente entre duas variáveis aleatórias independentes em que o numerador é uma variável aleatória com distribuição  $N(0, 1)$  e o denominador é a raiz quadrada de uma variável aleatória com distribuição quiquadrado com  $n-1$  graus de liberdade (veja (ii)) dividido pelo número de graus de liberdade, de modo que a variável (5.1.3) tem distribuição  $t$  de Student com  $n-1$  graus de liberdade.

## 5.2 O Método da Quantidade Pivotal

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

**Definição 5.2.1.** *Uma variável aleatória  $Q(X_1, \dots, X_n; \theta) = Q(\mathbf{X}; \theta)$  é dita ser uma quantidade pivotal para o parâmetro  $\theta$  se sua distribuição for independente de  $\theta$ .*

Notemos que uma quantidade pivotal não é uma estatística, pois ela depende de um parâmetro  $\theta$  desconhecido. Podemos, então, para cada  $\gamma = 1 - \alpha$  fixado, encontrar  $\lambda_1$  e  $\lambda_2$  na distribuição de  $Q(\mathbf{X}; \theta)$  de modo que

$$(5.2.1) \quad P[\lambda_1 \leq Q(\mathbf{X}; \theta) \leq \lambda_2] = \gamma.$$

Sendo a distribuição de  $Q(\mathbf{X}; \theta)$  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

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

e então de (5.2.1),

$$(5.2.2) \quad P[t_1(\mathbf{X}) \leq \theta \leq t_2(\mathbf{X})] = \gamma,$$

de modo que  $[t_1(\mathbf{X}); t_2(\mathbf{X})]$  é um intervalo (aleatório) que contém  $\theta$  com probabilidade (coeficiente de confiança)  $\gamma = 1 - \alpha$ . Nos casos em que a distribuição da variável aleatória  $X$  é discreta, em geral, não se consegue determinar  $\lambda_1$  e  $\lambda_2$  de tal forma que (5.2.1) esteja satisfeita exatamente. Em tais casos, podemos escolher  $\lambda_1$  e  $\lambda_2$  tal que (5.2.1) esteja satisfeita para um coeficiente de confiança maior ou igual a  $\gamma$  (o mais próximo possível). Quando  $n$  é razoavelmente grande, uma alternativa seria considerar os intervalos de confiança baseados na distribuição do estimador de máxima verossimilhança que consideramos na Seção 3.5. Um outro ponto a salientar é que, na maioria dos casos, existem muitos pares  $(\lambda_1, \lambda_2)$  satisfazendo (5.2.1). 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 5.2.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da distribuição da variável aleatória  $X$ , com densidade

$$(5.2.3) \quad f(x|\theta) = \theta e^{-\theta x}, \quad \theta > 0, \quad x > 0.$$

Como vimos no Capítulo 2, a estatística  $T = \sum_{i=1}^n X_i$  é suficiente para  $\theta$ . Mas, como a distribuição de  $T$  é  $Gama(n; \theta)$ , temos que  $T$  não é uma quantidade pivotal para  $\theta$ . Por outro lado, a densidade de  $Q(\mathbf{X}; \theta) = 2\theta \sum_{i=1}^n X_i$  é dada por

$$(5.2.4) \quad f_Q(y) = \frac{y^{n-1} e^{-y/2}}{2^n \Gamma[n]}, \quad y > 0$$

que corresponde a densidade de uma distribuição quiquadrado com  $2n$  graus de liberdade, que denotamos por  $\chi_{2n}^2$ . Portanto  $Q(\mathbf{X}; \theta)$  pode ser considerada como uma quantidade pivotal, pois sua distribuição é independente de  $\theta$ . Então, dado o coeficiente de confiança  $\gamma = 1 - \alpha$ , obtemos  $\lambda_1$  e  $\lambda_2$  na tabela da distribuição  $\chi_{2n}^2$ , de modo que

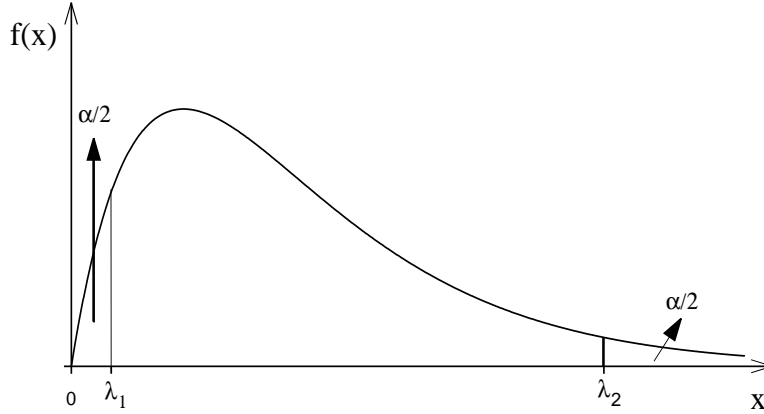
$$(5.2.5) \quad P \left[ \lambda_1 \leq 2\theta \sum_{i=1}^n X_i \leq \lambda_2 \right] = \gamma,$$

logo um intervalo de confiança para  $\theta$  com coeficiente de confiança  $\gamma$  é dado por

$$(5.2.6) \quad \left[ \frac{\lambda_1}{2 \sum_{i=1}^n X_i}; \frac{\lambda_2}{2 \sum_{i=1}^n X_i} \right].$$

Conforme enfatizado anteriormente, existem infinitos pares  $(\lambda_1, \lambda_2)$  para os quais (5.2.5) está verificada. Sempre que possível,  $(\lambda_1, \lambda_2)$  devem ser escolhidos de modo que o intervalo (5.2.6) seja de comprimento mínimo. Tal intervalo existe, mas  $(\lambda_1, \lambda_2)$  deve ser obtido por métodos computacionais. Uma alternativa é considerarmos intervalos simétricos em que  $(\lambda_1, \lambda_2)$  são obtidos a partir da distribuição  $\chi^2_{2n}$ , de modo que a área à esquerda de  $\lambda_1$  seja igual à área à direita de  $\lambda_2$  e igual a  $\alpha/2$ . Ver Figura 5.1.

**Figura 5.1.** Determinação de  $\lambda_1$  e  $\lambda_2$



Denotando estes pontos por  $q_1$  e  $q_2$ , temos que o intervalo simétrico é dado por

$$(5.2.7) \quad \left[ \frac{q_1}{2 \sum_{i=1}^n X_i}; \frac{q_2}{2 \sum_{i=1}^n X_i} \right].$$

A não ser que o tamanho da amostra  $n$  seja muito pequeno, o intervalo (5.2.7) é bastante próximo do intervalo de comprimento mínimo. Consideramos a seguir  $n = 20$  observações simuladas a partir da distribuição exponencial com  $\theta = 2$ . Como

$$F(x) = 1 - e^{-\theta x}$$

e como qualquer que seja a função de distribuição  $F(x)$

$$U = F(X) \sim U(0, 1),$$

ou seja, a distribuição de  $F(X)$  é uniforme no intervalo  $(0, 1)$ , gerando observações  $u$  a partir da distribuição  $U(0, 1)$ , temos que

$$(5.2.8) \quad x = -\frac{1}{\theta} \log(1 - u)$$

é uma observação simulada da distribuição exponencial com parâmetro  $\theta$  e com densidade dada em (5.2.3). As  $n = 20$  observações simuladas a partir da  $U(0, 1)$  são dadas na Tabela 5.1 abaixo.

**Tabela 5.1.**  $n = 20$  observações da  $U(0, 1)$

0,659	0,591	0,381	0,658	0,012
0,469	0,017	0,128	0,328	0,166
0,353	0,594	0,051	0,757	0,045
0,847	0,749	0,535	0,700	0,781

Usando os valores da Tabela 5.1 na relação (5.2.8) temos na Tabela 5.2 as  $n = 20$  observações simuladas da distribuição exponencial (5.2.3) com  $\theta = 2$ .

**Tabela 5.2.**  $n = 20$  observações da distribuição  $Exp(2)$

0,5380	0,4470	0,2398	0,5365	0,0061
0,3165	0,0086	0,0064	0,1995	0,9008
0,2177	0,4507	0,0262	0,7073	0,0230
0,9339	0,6912	0,3829	0,6020	0,7593

Considerando as primeiras  $n = 10$  observações na Tabela 5.2, temos que  $\sum_{i=1}^{10} X_i = 3,1992$ . Tomando  $\alpha = 0,05$ , temos da tabela da distribuição qui-quadrado com 20 graus de liberdade que  $q_1 = 9,59$  e  $q_2 = 34,17$ , então de (5.2.7) segue que o intervalo  $[1, 50; 5, 34]$  é um intervalo de confiança para  $\theta$  com coeficiente de confiança  $\gamma = 0,95$ . Considerando  $n = 20$ , temos que  $\sum_{i=1}^{20} X_i = 7,9934$  e usando a aproximação normal para a distribuição qui-quadrado (a maioria das tabelas da distribuição qui-quadrado não trazem percentis para 40 graus de liberdade), ou seja,

$$\frac{\chi_{2n}^2 - E[\chi_{2n}^2]}{\sqrt{Var[\chi_{2n}^2]}} \stackrel{a}{\sim} N(0, 1)$$

temos, usando a tabela da distribuição  $N(0, 1)$ , que

$$q_1 = -1,96\sqrt{80} + 40 \quad \text{e} \quad q_2 = 1,96\sqrt{80} + 40,$$

de modo que, nesse caso, o intervalo é dado por  $[1, 41; 3, 60]$  que, conforme era esperado, tem comprimento bem menor que o comprimento do correspondente intervalo com  $n = 10$ .

**Exemplo 5.2.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$  com distribuição uniforme no intervalo  $(0, \theta)$ , ou seja,  $X \sim U(0, \theta)$ . Vimos no Capítulo 2 que uma estatística suficiente para  $\theta$  é dada por  $Y = X_{(n)} = \max\{X_1, \dots, X_n\}$ , com função de densidade dada por

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

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

$$(5.2.9) \quad f_Q(q) = nq^{n-1} I_{[0, 1]}(q)$$

que não depende de  $\theta$ . Portanto a variável aleatória  $Q(\mathbf{X}; \theta)$  é uma quantidade pivotal, de modo que dado  $\gamma = 1 - \alpha$ , podemos encontrar  $\lambda_1$  e  $\lambda_2$  na distribuição de  $Q$ , tal que

$$(5.2.10) \quad \int_{\lambda_1}^{\lambda_2} f_Q(q) dq = \gamma = 1 - \alpha.$$

Como existem infinitos pares  $(\lambda_1, \lambda_2)$  satisfazendo (5.2.10), consideramos o intervalo simétrico, ou seja, consideramos o intervalo satisfazendo

$$(5.2.11) \quad \int_0^{\lambda_1} f_Q(q) dq = \frac{\alpha}{2} \quad \text{e} \quad \int_{\lambda_2}^1 f_Q(q) dq = \frac{\alpha}{2}.$$

Resolvendo as equações (5.2.11), chegamos a

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

de modo que

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

que leva ao intervalo

$$(5.2.12) \quad \left[ \frac{X_{(n)}}{(1 - \alpha/2)^{1/n}}; \frac{X_{(n)}}{(\alpha/2)^{1/n}} \right].$$

Considerando as primeiras  $n = 10$  observações da Tabela 5.1 e  $\gamma = 0,95$ , temos que o intervalo (5.2.12) se reduz a  $[0,659/(0,975)^{1/10}; 0,659/(0,025)^{1/10}]$ , ou seja,  $[0,661; 0,953]$ . Considerando as  $n = 20$  observações da Tabela 5.1, o intervalo se reduz a  $(0,848; 1,019)$ . Notemos que  $\theta = 1$  não está contido no intervalo com  $n = 10$ , mas está contido no intervalo com  $n = 20$ . Como a distribuição de  $Q$  não é simétrica, o intervalo (5.2.12) não é o de menor comprimento para um dado  $\gamma$ . No Exercício 5.3 apresentamos um intervalo de menor comprimento que o do intervalo (5.2.12). É importante ressaltar que o coeficiente de confiança  $\gamma$  está associado ao intervalo aleatório que segue de (5.2.2). Quanto ao intervalo numérico que segue do intervalo aleatório, afirmações do tipo  $P[0,848 \leq \theta \leq 1,019]$  não são apropriadas, pois não existem quantidades aleatórias associadas à desigualdade  $0,848 \leq \theta \leq 1,019$ . O que se aplica no caso numérico é a interpretação freqüentista, ou seja, para cada 100 intervalos numéricos construídos a partir do intervalo aleatório, aproximadamente  $100\gamma\%$  deles vão conter  $\theta$ . Para um problema particular, o intervalo que construímos a partir de uma amostra observada pode ser ou não um daqueles  $100(1 - \gamma)\%$  que não contém  $\theta$ . Mas não temos condições de sabê-lo.

### 5.3 Intervalos para Populações Normais

Consideremos em primeiro lugar (Seção 5.3.1) o caso de uma única amostra. A seguir, na Seção 5.3.2, abordamos o caso de duas amostras.

#### 5.3.1 O caso de uma única amostra

Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição  $N(\mu, \sigma^2)$ . Assumindo  $\sigma^2$  conhecido, temos que uma quantidade pivotal para  $\mu$ , baseada na estatística suficiente  $\sum_{i=1}^n X_i = n\bar{X}$  é dada por

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

que tem distribuição  $N(0, 1)$ . Portanto, dado o coeficiente de confiança  $\gamma$ , determinamos  $\lambda_1$  e  $\lambda_2$  de modo que

$$(5.3.1) \quad P \left[ \lambda_1 \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \lambda_2 \right] = \gamma.$$

Conforme enfatizado anteriormente, existem infinitos pares  $(\lambda_1, \lambda_2)$  que satisfazem (5.3.1). Como a distribuição  $N(0, 1)$  é simétrica, o intervalo de menor comprimento é o intervalo simétrico, ou seja, aquele em que a área à direita de  $\lambda_2$  é igual a área à esquerda de  $\lambda_1$  que é igual a  $\alpha/2$ . Sejam então  $\lambda_1 = -z_{\alpha/2}$  e

$\lambda_2 = z_{\alpha/2}$ , onde  $P(Z \leq z_{\alpha/2}) = 1 - \alpha/2$ ,  $Z \sim N(0, 1)$  de modo que o intervalo de menor comprimento é dado por

$$(5.3.2) \quad \left[ \bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right].$$

Por outro lado, sendo  $\sigma^2$  desconhecido, temos pelo Teorema 5.1. (iii), que

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

que nesse caso é uma quantidade pivotal. Então, dado  $\gamma$ , existem  $\lambda_1$  e  $\lambda_2$  na distribuição  $t_{n-1}$  de modo que

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

Como a distribuição da quantidade pivotal  $Q$  é simétrica, devemos escolher  $\lambda_1$  e  $\lambda_2$  de modo que a área à direita de  $\lambda_2$  seja igual a área à esquerda de  $\lambda_1$ , ou seja  $\lambda_1 = -t_{\alpha/2}$  e  $\lambda_2 = t_{\alpha/2}$ , onde  $P(T \leq t_{\alpha/2}) = 1 - \alpha/2$ ,  $T \sim t_{n-1}$  de modo que o intervalo de menor comprimento é dado por

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

Quanto a  $\sigma^2$ , considerando  $\mu$  desconhecido, temos, de acordo com o Teorema 5.1. (ii), que

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

é uma quantidade pivotal para  $\sigma^2$ . Portanto, dado  $\gamma$ , podemos determinar  $\lambda_1$  e  $\lambda_2$  de modo que

$$(5.3.3) \quad P \left[ \lambda_1 \leq \frac{(n-1)S^2}{\sigma^2} \leq \lambda_2 \right] = \gamma.$$

Considerando o intervalo simétrico, ou seja,  $\lambda_1 = q_1$  e  $\lambda_2 = q_2$ , onde  $P[\chi_{n-1}^2 \geq q_2] = P[\chi_{n-1}^2 \leq q_1] = \alpha/2$ , temos de (5.3.3), o intervalo

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

### 5.3.2 Duas amostras independentes

Vamos considerar o caso em que temos  $X_1, \dots, X_n$ , uma amostra aleatória da variável aleatória  $X \sim N(\mu_1, \sigma^2)$  e  $Y_1, \dots, Y_m$ , uma amostra aleatória da variável aleatória  $Y \sim N(\mu_2, \sigma^2)$ , onde  $X$  e  $Y$  são independentes. Sabemos que

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

de modo que, sendo  $\theta = \mu_1 - \mu_2$ , consideramos a quantidade pivotal

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

Sendo  $\sigma^2$  conhecido, temos, como na seção anterior, o intervalo

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

onde  $z_{\alpha/2}$  é obtido como em (5.3.2). Sendo  $\sigma^2$  desconhecido, temos que uma quantidade pivotal é dada por

$$(5.3.4) \quad Q(\mathbf{X}, \mathbf{Y}, \theta) = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$$

onde

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{(n+m-2)}, \quad S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

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

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

$$(5.3.5) \quad \frac{(n+m-2)S_p^2}{\sigma^2} = \frac{(n-1)S_x^2 + (m-1)S_y^2}{\sigma^2} \sim \chi_{n+m-2}^2.$$

Então do Teorema 5.1, (iii) segue o resultado (5.3.4). Um intervalo de confiança para  $\theta = \mu_1 - \mu_2$ , com coeficiente de confiança  $\gamma$  é, então, dado por



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

onde  $t_{\alpha/2}$  é obtido na tabela da distribuição  $t$  com  $n+m-2$  graus de liberdade. Para construirmos um intervalo de confiança para  $\sigma^2$ , podemos considerar a quantidade pivotal (5.3.5).

No caso em que  $X \sim N(\mu_1, \sigma_1^2)$  e  $Y \sim N(\mu_2, \sigma_2^2)$  e o interesse é a construção de um intervalo de confiança para  $\sigma_1^2/\sigma_2^2$ , notando que

$$\frac{(n-1)S_x^2}{\sigma_1^2} \sim \chi_{n-1}^2 \quad \text{e} \quad \frac{(m-1)S_y^2}{\sigma_2^2} \sim \chi_{m-1}^2,$$

temos que

$$Q(\mathbf{X}, \mathbf{Y}, \theta) = \frac{(m-1)S_y^2/\sigma_2^2(m-1)}{(n-1)S_x^2/\sigma_1^2(n-1)} \sim F_{m-1, n-1},$$

onde  $F_{m-1, n-1}$  denota a distribuição  $F$  com  $m-1$  e  $n-1$  graus de liberdade, é uma quantidade pivotal para  $\theta$ . Então, dado  $\gamma$ , obtemos  $\lambda_1$  e  $\lambda_2$  na distribuição  $F_{m-1, n-1}$ , de modo que

$$P \left[ \lambda_1 \leq \frac{\sigma_1^2 S_y^2}{\sigma_2^2 S_x^2} \leq \lambda_2 \right] = \gamma$$

Considerando o intervalo simétrico, ou seja,  $\lambda_1 = F_1$  e  $\lambda_2 = F_2$ , de modo que

$$P[F_{m-1, n-1} \geq F_2] = P[F_{m-1, n-1} \leq F_1] = \alpha/2,$$

onde  $F_1$  e  $F_2$  são obtidos na tabela da distribuição  $F$  com  $m-1$  e  $n-1$  graus de liberdade, temos o intervalo

$$\left[ F_1 \frac{S_x^2}{S_y^2}; F_2 \frac{S_x^2}{S_y^2} \right].$$

## 5.4 Intervalos de Confiança Aproximados

Nesta seção consideramos intervalos de confiança aproximados para um parâmetro  $\theta$  baseados na distribuição assintótica do estimador de máxima verossimilhança  $\hat{\theta}$  de  $\theta$ . De acordo com (3.2.3), temos que

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

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

$$(5.4.1) \quad Q(\mathbf{X}, \theta) = \frac{\hat{\theta} - \theta}{\sqrt{(nI_F(\hat{\theta}))^{-1}}} \stackrel{a}{\sim} N(0, 1),$$

de modo que  $Q(\mathbf{X}, \theta)$  é uma quantidade pivotal com distribuição aproximadamente igual a distribuição  $N(0, 1)$  em grandes amostras. Com relação a uma função  $g(\theta)$ , podemos considerar a variável aleatória

$$(5.4.2) \quad Q(\mathbf{X}, g(\theta)) = \frac{g(\hat{\theta}) - g(\theta)}{\sqrt{\frac{(g'(\hat{\theta}))^2}{nI_F(\hat{\theta})}}} \stackrel{a}{\sim} N(0, 1),$$

que para amostras grandes é uma quantidade pivotal.

**Exemplo 5.4.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim \text{Bernoulli}(\theta)$ . Como o estimador de máxima verossimilhança de  $\theta$  é  $\hat{\theta} = \bar{X}$  e  $I_F(\theta) = 1/\theta(1 - \theta)$ , de (5.4.1), temos que uma quantidade pivotal para  $\theta$  é dada por

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

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

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

Suponhamos agora, que seja de interesse a obtenção de um intervalo de confiança para  $g(\theta) = \theta(1 - \theta)$ . Como  $g'(\theta) = 1 - 2\theta$  e  $I_F(\theta) = 1/\theta(1 - \theta)$ , temos de (5.4.2) que uma quantidade pivotal para  $g(\theta)$  é dada por

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

de modo que um intervalo de confiança aproximado para  $g(\theta) = \theta(1 - \theta)$  é dado por

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

onde  $z_{\alpha/2}$  é obtido na tabela da distribuição  $N(0, 1)$ .

**Exemplo 5.4.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X \sim \text{Exp}(\theta)$ , com função densidade

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

Como  $I_F^{-1}(\theta) = \theta^2$  e  $\hat{\theta} = 1/\bar{X}$ , segue de (5.4.1) que uma quantidade pivotal para  $\theta$  é dada por

$$Q(\mathbf{X}, \theta) = \frac{1/\bar{X} - \theta}{\sqrt{\hat{\theta}^2/n}} \stackrel{a}{\sim} N(0, 1),$$

de modo que um intervalo de confiança com coeficiente de confiança aproximado  $\gamma = 1 - \alpha$  é dado por

$$(5.4.3) \quad \left[ \frac{1}{\bar{X}} - z_{\alpha/2} \sqrt{\frac{1}{n\bar{X}^2}}; \frac{1}{\bar{X}} + z_{\alpha/2} \sqrt{\frac{1}{n\bar{X}^2}} \right].$$

Considerando a amostra da Tabela 5.2, temos que para  $n = 10$  o intervalo (5.4.3) se reduz a (1,189;5,063) e para  $n = 20$ , temos o intervalo (1,405;3,599). Notemos que o intervalo aproximado para  $\theta$  com  $n = 20$  coincide com o intervalo exato obtido no Exemplo 5.2.1.

## 5.5 Intervalos de Confiança Bayesianos

Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$  com função densidade de probabilidade (ou função de probabilidade)  $f(x|\theta)$ . Consideremos para  $\theta$  a função de densidade a priori  $\pi(\theta)$ . Portanto a função de densidade a posteriori para  $\theta$ , é, de acordo com (4.4.6), dada por

$$\pi(\theta|\mathbf{X}) = \frac{\prod_{i=1}^n f(x_i|\theta)\pi(\theta)}{\int_{\Theta} \prod_{i=1}^n f(x_i|\theta)\pi(\theta)d\theta}.$$

**Definição 5.5.1.** Dizemos que  $[t_1; t_2]$  é um intervalo de confiança Bayesiano para  $\theta$ , com coeficiente de confiança  $\gamma = 1 - \alpha$  se

$$(5.5.1) \quad \int_{t_1}^{t_2} \pi(\theta|\mathbf{X})d\theta = \gamma.$$

Como no caso clássico existem, em geral, infinitos intervalos  $[t_1; t_2]$  satisfazendo (5.5.1). Sempre que possível, o comprimento do intervalo  $[t_1; t_2]$  deve ser mínimo. Nos casos em que a função de densidade a posteriori é simétrica,

os intervalos simétricos so em geral os de menor comprimento. O intervalo Bayesiano de menor comprimento é usualmente conhecido como o intervalo de densidade a posteriori mxima “highest posterior density (HPD) interval”. Métodos computacionais so em geral necessrios para a obtenço do intervalo HPD.

**Exemplo 5.5.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatria de tamanho  $n$  da distribuiço  $N(\mu, 1)$ . Consideremos para  $\mu$  a distribuiço a priori  $N(\mu_0, 1)$ . Do Exemplo 4.4.3, temos que a distribuiço a posteriori de  $\mu$  dado  $\mathbf{X}$  que denotamos por  $\mu|\mathbf{X}$ , é dada por

$$\mu|\mathbf{X} \sim N\left(\frac{\sum_{i=1}^n X_i + \mu_0}{n+1}, \frac{1}{n+1}\right).$$

Sendo  $\gamma = 0,95$ , temos ento de (5.5.1) e da tabela da distribuiço  $N(0, 1)$  que  $[t_1; t_2]$  deve ser escolhido de modo que

$$\frac{t_1 - \frac{\sum_{i=1}^n X_i + \mu_0}{n+1}}{\sqrt{\frac{1}{n+1}}} = -1,96 \quad \text{e} \quad \frac{t_2 - \frac{\sum_{i=1}^n X_i + \mu_0}{n+1}}{\sqrt{\frac{1}{n+1}}} = 1,96,$$

ou seja,

$$t_1 = \frac{\sum_{i=1}^n X_i + \mu_0}{n+1} - 1,96\sqrt{\frac{1}{n+1}} \quad \text{e} \quad t_2 = \frac{\sum_{i=1}^n X_i + \mu_0}{n+1} + 1,96\sqrt{\frac{1}{n+1}},$$

logo o intervalo Bayesiano de menor comprimento (HPD) para  $\mu$  com coeficiente de confiança  $\gamma = 0,95$  é dado por

$$\left[ \frac{\sum_{i=1}^n X_i + \mu_0}{n+1} - 1,96\sqrt{\frac{1}{n+1}}; \frac{\sum_{i=1}^n X_i + \mu_0}{n+1} + 1,96\sqrt{\frac{1}{n+1}} \right].$$

**Exemplo 5.5.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatria de tamanho  $n$  da varivel aleatria  $X \sim U(0, \theta)$ . Consideremos para  $\theta$  a priori com densidade (Pareto)

$$\pi(\theta) = \frac{ba^b}{\theta^{b+1}} I_{(a, \infty)}(\theta).$$

Do Exerccio 4.10, temos que a densidade a posteriori de  $\theta$  dado  $X_1, \dots, X_n$  é dada por

$$(5.5.2) \quad h(\theta|\mathbf{X}) = \frac{(n+b)(\max(a, X_{(n)}))^{n+b}}{\theta^{n+b+1}} I_{(\max(a, X_{(n)}), \infty)}(\theta).$$

Ento, temos de (5.5.1) que o intervalo Bayesiano “simétrico” para  $\theta$ , com coeficiente de confiança  $\gamma = 1 - \alpha$  é obtido pela soluço das equaçes

$$\int_{\max(a, X_{(n)})}^{t_1} \frac{(n+b)\max(a, X_{(n)})^{n+b}}{\theta^{n+b+1}} d\theta = \frac{\alpha}{2}$$

e

$$\int_{t_2}^{\infty} \frac{(n+b)\max(a, X_{(n)})^{n+b}}{\theta^{n+b+1}} d\theta = \frac{\alpha}{2},$$

o que leva a

$$t_1 = \frac{\max(a, X_{(n)})}{(1-\alpha/2)^{1/n+b}} \quad \text{e} \quad t_2 = \frac{\max(a, X_{(n)})}{(\alpha/2)^{1/n+b}},$$

de modo que o intervalo Bayesiano simétrico para  $\theta$ , com coeficiente de confiança  $\gamma = 1 - \alpha$ , é dado por

$$(5.5.3) \quad \left[ \frac{\max(a, X_{(n)})}{(1-\alpha/2)^{1/n+b}}; \frac{\max(a, X_{(n)})}{\alpha/2^{1/n+b}} \right].$$

Desde que a densidade a posteriori (5.5.2) não é simétrica, temos que o intervalo (5.5.3) não é o HPD que nesse caso deve ser obtido numericamente.

## 5.6 Exercícios

**5.1.** Verifique a validade da expressão (5.1.1).

**5.2.** Considere o Exemplo 5.2.1. Mostre que a distribuição da quantidade pivotal

$$Q(\mathbf{X}, \theta) = 2\theta \sum_{i=1}^n X_i$$

é quiquadrado com  $2n$  graus de liberdade com densidade dada por (5.2.4).

**5.3.** Considere o Exemplo 5.2.2. Mostre que a distribuição de  $Q(\mathbf{X}, \theta) = X_{(n)}/\theta$  é dada por (5.2.9). Considere o intervalo

$$(5.6.1) \quad \left[ X_{(n)}; \frac{X_{(n)}}{\alpha^{1/n}} \right].$$

Encontre seu coeficiente de confiança, compare seu comprimento com o do intervalo obtido no Exemplo 5.2.2, e mostre que o intervalo (5.6.1) é o de menor comprimento dentre todos os intervalos com coeficiente de confiança  $\gamma = 1 - \alpha$ .

**5.4.** Seja  $X$  uma única observação da densidade

$$f(x|\theta) = \theta x^{\theta-1} \quad 0 < x < 1, \quad \theta > 0.$$

- (i) Mostre que  $-\theta \log X$  é uma quantidade pivotal e use-a para construir um intervalo de confiança para  $\theta$  com coeficiente de confiança  $\gamma = 1 - \alpha$ .  
 (ii) Seja  $Y = (-\log X)^{-1}$ . Encontre o coeficiente de confiança associado ao intervalo  $(Y/2, Y)$ .

**5.5.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\theta, \theta)$ . Sugira uma quantidade pivotal para construir um intervalo de confiança para  $\theta$  com  $\gamma = 1 - \alpha$ .

**5.6.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com função de densidade de probabilidade dada por

$$f(x|\theta) = I_{(\theta-1/2, \theta+1/2)}(x).$$

Seja  $[X_{(1)}; X_{(n)}]$  um intervalo de confiança para  $\theta$ . Calcule seu coeficiente de confiança. Mostre que o resultado vale para qualquer distribuição simétrica em torno de  $\theta$ .

**5.7.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com função densidade de probabilidade dada por

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

Encontre intervalos de confiança para  $E(X)$  e  $Var(X)$  com coeficientes de confiança  $\gamma = 1 - \alpha$ .

**5.8.** Sejam  $X_1, X_2$  uma amostra aleatória de tamanho 2 da distribuição  $N(\mu, 1)$ . Seja  $Y_1 < Y_2$  a amostra ordenada correspondente.

- (i) Encontre o coeficiente de confiança associado ao intervalo  $(Y_1, Y_2)$ .  
 (ii) Considere o intervalo de confiança para  $\mu$  baseado na quantidade pivotal  $\bar{X} - \mu$ , onde  $\bar{X} = (X_1 + X_2)/2$ . Compare o comprimento esperado deste intervalo com o comprimento esperado do intervalo em (i) usando o mesmo  $\gamma$ .

**5.9.** Sejam  $X_1, \dots, X_{n+1}$ , uma amostra aleatória de tamanho  $n+1$  ( $n > 1$ ) da distribuição  $N(\mu, \sigma^2)$ , onde  $\mu$  e  $\sigma^2$  são desconhecidos.

- (i) Encontre  $c$  tal que

$$\frac{c(\bar{X} - X_{n+1})}{S} \sim t_{n-1},$$

onde

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{e} \quad S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

- (ii) Se  $n = 8$ , encontre  $k$  de modo que

$$P[\bar{X} - kS \leq X_9 \leq \bar{X} + kS] = 0,80.$$

**5.10.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim \text{Exp}(\theta_1)$  e  $Y_1, \dots, Y_m$  uma amostra aleatória da variável aleatória  $Y \sim \text{Exp}(\theta_2)$ . Assumindo que as duas amostras são independentes,

- (i) obtenha uma quantidade pivotal para construir um intervalo de confiança para  $\theta_1/\theta_2$ .
- (ii) Suponha que  $\theta_1 = 1,5$  e  $\theta_2 = 2,0$ . Simule uma amostra aleatória com  $n = 10$  da variável  $X$  e com  $m = 15$  da variável aleatória  $Y$ . Como fica o seu intervalo obtido a partir da quantidade pivotal encontrada em (i)?

**5.11.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição  $\text{Poisson}(\theta)$ , com priori

$$\pi(\theta) = e^{-\theta}, \quad \theta > 0.$$

Construa um intervalo de confiança Bayesiano simétrico para  $\theta$  com  $\gamma = 0,95$ . Se  $n = 10$  e  $\sum_{i=1}^n X_i = 18$ , como fica o intervalo?

**5.12.** Considere o Exercício 4.9. Obtenha um intervalo de confiança Bayesiano para  $\theta$  com coeficiente de confiança  $\gamma = 0,95$ . Como fica seu intervalo se  $x = 4$ ?

**5.13.** Considere o Exercício 4.12. Construa um intervalo de confiança para  $\theta$  com coeficiente de confiança  $\gamma = 1 - \alpha$ , sendo  $r = \lambda = 2$ . Considere  $\theta = 2$  e simule uma amostra de  $X$  com  $n = 10$ . Como fica o intervalo com  $\gamma = 0,95$ ?

**5.14.** Usando a amostra de tamanho  $n = 20$  no Exemplo 3.1.6, construa um intervalo aproximado para  $\theta$ , onde  $f(x|\theta)$  é dada em (3.1.8).





## 6. Testes de Hipóteses

Neste capítulo apresentamos a teoria de testes de hipóteses em um nível bastante introdutório. Testes ótimos, como os testes mais poderosos para hipótese nula simples contra alternativa simples e testes uniformemente mais poderosos para hipóteses compostas, são obtidos utilizando o conhecido Lema de Neyman-Pearson. Situações mais complexas, como o caso de hipóteses bilaterais, são tratadas utilizando-se a estatística da razão de verossimilhanças generalizada que, apesar de não apresentar propriedades ótimas, tem um comportamento bastante satisfatório.

### 6.1 Idéias Básicas

Em muitas situações temos interesse em tomar a decisão de aceitar ou rejeitar determinada afirmação baseando-se em um conjunto de evidências. Um exemplo comum é o caso em que um indivíduo está sendo julgado por determinado delito. Com base nas evidências (testemunhas, fatos, etc.), o júri terá que decidir pela culpa ou inocência do indivíduo. Podemos, então, concluir que o júri formula duas hipóteses: “ $H_0$  : o indivíduo é inocente” e a alternativa “ $H_1$  : o indivíduo é culpado”. Com base nas evidências apresentadas, o júri terá que se decidir por  $H_0$  ou por  $H_1$ . Ao tomar, por exemplo, a decisão de aceitar  $H_1$  (então rejeitar  $H_0$ ) como verdadeira, o júri pode estar cometendo um erro, pois, apesar das evidências, o indivíduo pode ser inocente. O mesmo pode acontecer com relação à aceitação da hipótese  $H_0$  como verdadeira. Nesse caso, o júri estaria considerando como inocente um indivíduo culpado.

Um problema mais próximo da área de atuação da estatística (apesar de que muita estatística tem sido utilizada em problemas jurídicos), é o problema de se decidir sobre a eficiência ou não de certa vacina utilizada no combate à determinada doença. Os pesquisadores formulam então as hipóteses “ $H_0$  : a vacina não é eficiente” e “ $H_1$  : a vacina é eficiente”. Nesse caso, um experimento é planejado, envolvendo um grupo possivelmente grande de indivíduos em que uma parte (escolhida ao acaso) recebe a vacina e o restante recebe uma substância inóqua. Com base nos resultados desse experimento, os pesquisadores terão

então que se decidir por  $H_0$  ou  $H_1$ . Novamente, não está descartada a possibilidade de que erros sejam cometidos ao se considerar, por exemplo, a vacina eficiente ( $H_0$  falsa) quando, na verdade, ela não o é ( $H_0$  é verdadeira), o que seria bastante prejudicial à população. O estatístico envolvido na pesquisa deve procurar utilizar técnicas que tornem mínima a probabilidade de se cometer erros.

## 6.2 Formulação Estatística

Nesta seção os princípios básicos da teoria são especificados. Formalizamos a seguir a noção de hipótese estatística.

**Definição 6.2.1.** *Chamamos de hipótese estatística qualquer afirmação acerca da distribuição de probabilidades de uma ou mais variáveis aleatórias.*

Denotamos por  $H_0$  (hipótese nula) a hipótese de interesse. Caso  $H_0$  seja rejeitada, aceitamos como verdadeira a hipótese alternativa  $H_1$ . Sendo a variável aleatória  $X$  distribuída de acordo com a função de densidade (ou de probabilidade)  $f(x|\theta)$ , com  $\theta \in \Theta$ , dizemos que a distribuição de  $X$  está totalmente especificada quando conhecemos  $f(x|\theta)$  e  $\theta$ . A distribuição de  $X$  será dita estar parcialmente especificada quando conhecemos a função de densidade (ou de probabilidade)  $f(x|\theta)$ , mas não  $\theta$ . Associados às hipóteses  $H_0$  e  $H_1$ , definimos os conjuntos  $\Theta_0$  e  $\Theta_1$ , ou seja,  $H_0$  afirma que  $\theta \in \Theta_0$  (notação:  $H_0 : \theta \in \Theta_0$ ) e  $H_1$  afirma que  $\theta \in \Theta_1$  (notação:  $H_1 : \theta \in \Theta_1$ ). No caso em que  $\Theta_0 = \{\theta_0\}$  dizemos que  $H_0$  é simples. Caso contrário, dizemos que  $H_0$  é composta. O mesmo vale para a hipótese alternativa  $H_1$ .

**Definição 6.2.2.** *Chamamos de teste de uma hipótese estatística a função de decisão  $d : \mathcal{X} \rightarrow \{a_0, a_1\}$ , em que  $a_0$  corresponde à ação de considerar a hipótese  $H_0$  como verdadeira e  $a_1$  corresponde à ação de considerar a hipótese  $H_1$  como verdadeira.*

Na definição acima,  $\mathcal{X}$  denota o espaço amostral associado à amostra  $X_1, \dots, X_n$ . A função de decisão  $d$  divide o espaço amostral  $\mathcal{X}$  em dois conjuntos

$$\mathbf{A}_0 = \{(x_1, \dots, x_n) \in \mathcal{X}; d(x_1, \dots, x_n) = a_0\}$$

e

$$\mathbf{A}_1 = \{(x_1, \dots, x_n) \in \mathcal{X}; d(x_1, \dots, x_n) = a_1\},$$

onde  $\mathbf{A}_0 \cup \mathbf{A}_1 = \mathcal{X}$  e  $\mathbf{A}_0 \cap \mathbf{A}_1 = \emptyset$ . Como em  $\mathbf{A}_0$  temos os pontos amostrais  $\mathbf{x} = (x_1, \dots, x_n)$  que levam à aceitação de  $H_0$ , vamos chamar  $\mathbf{A}_0$  de região de *aceitação* e, por analogia,  $\mathbf{A}_1$  de região de *rejeição* de  $H_0$ , também chamada de *região crítica*.

**Exemplo 6.2.1.** Uma caixa contém duas moedas. Uma apresenta cara com probabilidade  $p = 0,5$  (equilibrada) e a outra apresenta cara com probabilidade  $p = 0,6$ . Uma moeda é escolhida aleatoriamente e lançada três vezes. Suponhamos que as hipóteses de interesse são  $H_0 : p = 0,5$  e  $H_1 : p = 0,6$ . Seja  $X_i$  a variável de Bernoulli que assume o valor 1 se ocorre cara no  $i$ -ésimo lançamento e 0 caso contrário,  $i = 1, 2, 3$ . Nesse caso,

$$\mathcal{X} = \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (0, 1, 1), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}.$$

Podemos considerar, por exemplo, a região crítica

$$\mathbf{A}_1 = \{(x_1, x_2, x_3); x_1 + x_2 + x_3 \geq 2\},$$

de modo que

$$\mathbf{A}_0 = \{(x_1, x_2, x_3); x_1 + x_2 + x_3 < 2\}.$$

Notemos que  $\mathbf{A}_0 \cup \mathbf{A}_1 = \mathcal{X}$  e  $\mathbf{A}_0 \cap \mathbf{A}_1 = \emptyset$ .

No caso em que  $H_0 : \theta = \theta_0$  (simples) e  $H_1 : \theta = \theta_1$  (simples), considerando a função de perda  $l(\theta, d) = 0$  ou 1, se a decisão correta ou incorreta, respectivamente, é tomada, a função de risco é, então, dada por

$$\begin{aligned} R(\theta_0, d) &= E[l(\theta_0, d)] = 0.P[\mathbf{X} \in \mathbf{A}_0|\theta_0] + 1.P[\mathbf{X} \in \mathbf{A}_1|\theta_0] \\ &= P[\mathbf{X} \in \mathbf{A}_1|\theta_0] = \alpha = P_{H_0}[\text{Rejeitar } H_0] \end{aligned}$$

e

$$\begin{aligned} R(\theta_1, d) &= E[l(\theta_1, d)] = 0.P[\mathbf{X} \in \mathbf{A}_1|\theta_1] + 1.P[\mathbf{X} \in \mathbf{A}_0|\theta_1] \\ &= P[\mathbf{X} \in \mathbf{A}_0|\theta_1] = \beta = P_{H_1}[\text{aceitar } H_0]. \end{aligned}$$

Os riscos  $\alpha$  e  $\beta$  são conhecidos na literatura como probabilidades dos erros dos tipos I e II, respectivamente. Mais precisamente, o erro do tipo I ocorre quando rejeitamos  $H_0$ , sendo  $H_0$  verdadeira, enquanto que o erro do tipo II ocorre quando aceitamos  $H_0$ , sendo  $H_0$  falsa. A situação descrita acima está ilustrada na Tabela 6.1 dada abaixo.

**Tabela 6.1.** Tipos de erros em testes de hipóteses

Decisão	$H_0$ é verdadeira	$H_0$ é falsa
Aceitar $H_0$	Decisão correta	Erro do tipo II
Rejeitar $H_0$	Erro do tipo I	Decisão correta

**Definição 6.2.3.** O poder do teste com região crítica  $\mathbf{A}_1$  para testar  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta = \theta_1$  é dado por

$$(6.2.1) \quad \pi(\theta_1) = P_{H_1}[\mathbf{X} \in \mathbf{A}_1] = P[\mathbf{X} \in \mathbf{A}_1 | \theta_1].$$

Notemos de (6.2.1) que  $\pi(\theta_1) = 1 - \beta$ , onde  $\beta$  é a probabilidade de se cometer o erro do tipo II.

**Exemplo 6.2.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição da variável aleatória  $X \sim N(\mu, 1)$ . Consideremos as hipóteses  $H_0 : \mu = 0$  e  $H_1 : \mu = 1$ . Consideremos o teste com região crítica  $\mathbf{A}_1 = \{\mathbf{x}; \bar{x} \geq c\}$ , onde, como nos capítulos anteriores,  $\bar{x} = (x_1 + \dots + x_n)/n$ . Suponhamos que  $n = 16$  e que temos interesse em fixar  $\alpha = 0,05$ . Então, para determinar  $c$ , temos que resolver a equação  $\alpha = P_{H_0}[\bar{X} \geq c]$ , ou seja,

$$0,05 = P_{H_0}[\bar{X} \geq c] = P[Z \geq c\sqrt{n}],$$

onde  $Z = \bar{X}\sqrt{n} \sim N(0, 1)$ . Então,  $c\sqrt{n} = 1,64$ , pois na distribuição  $N(0, 1)$ , o valor 1,64 é o percentil 95%. Logo  $c = 0,41$ , de modo que  $\mathbf{A}_1 = \{\mathbf{x}, \bar{x} \geq 0,41\}$ .

### 6.3 Hipótese Nula Simples contra Alternativa Simples. Testes Mais Poderosos

Nesta seção, fixada a probabilidade do erro do tipo I,  $\alpha$ , também conhecida como nível do teste, procuramos a região crítica  $\mathbf{A}_1^*$  que tenha a menor probabilidade de erro do tipo II, ou seja, maior poder dentre todos os testes com nível menor ou igual a  $\alpha$ . Enfatizamos que, no caso discreto,

$$\alpha(\mathbf{A}_1) = P_{H_0}[\mathbf{X} \in \mathbf{A}_1] = \sum_{\mathbf{x} \in \mathbf{A}_1} f(\mathbf{x} | \theta_0) \quad \text{e} \quad \beta(\mathbf{A}_1) = \sum_{\mathbf{x} \in \mathbf{A}_0} f(\mathbf{x} | \theta_1),$$

onde  $\mathbf{A}_0 = \mathbf{A}_1^c$ , conforme enfatizado anteriormente.

**Exemplo 6.3.1.** Consideremos o problema de se testar  $H_0 : \theta = \theta_0$  versus  $H_1 : \theta = \theta_1$ , com uma única observação da variável aleatória  $X$ , com distribuição de probabilidade dada na Tabela 6.2 abaixo.

**Tabela 6.2.** Função de probabilidade da variável aleatória  $X$  sob  $H_0$  e  $H_1$

$X$	0	1	2	3	4	5
$f(x \theta_0)$	0,02	0,03	0,05	0,05	0,35	0,50
$f(x \theta_1)$	0,04	0,05	0,08	0,12	0,41	0,30

Notemos que as possíveis regiões críticas  $\mathbf{A}_1$  de nível  $\alpha(\mathbf{A}_1) = 0,05$  com os respectivos  $\beta = \beta(\mathbf{A}_1)$  são dadas na Tabela 6.3 abaixo.

**Tabela 6.3.** Regiões críticas  $\mathbf{A}_1$  com nível  $\alpha(\mathbf{A}_1) = 0,05$ 

$\mathbf{A}_1$	$\alpha$	$\mathbf{A}_0$	$\beta$
$\{0, 1\}$	0,05	$\{2, 3, 4, 5\}$	0,91
$\{2\}$	0,05	$\{0, 1, 3, 4, 5\}$	0,92
$\{3\}$	0,05	$\{0, 1, 2, 4, 5\}$	0,88

Portanto, dentre todas as regiões críticas de nível  $\alpha = 0,05$ , a mais poderosa (menor  $\beta$ ) é dada por  $\mathbf{A}_1 = \{3\}$ .

O resultado que segue apresenta o teste que minimiza uma combinação linear dos erros, do tipo  $a\alpha + b\beta$ , com  $a$  e  $b$  conhecidos.

**Lema 6.3.1.** *Consideremos o teste com região crítica*

$$\mathbf{A}_1^* = \left\{ \mathbf{x}; \frac{L_1(\mathbf{x})}{L_0(\mathbf{x})} \geq \frac{a}{b} \right\},$$

onde  $a$  e  $b$  são especificados e  $b > 0$ . Então, para qualquer outro teste com região crítica  $\mathbf{A}_1$ , temos que

$$a\alpha(\mathbf{A}_1^*) + b\beta(\mathbf{A}_1^*) \leq a\alpha(\mathbf{A}_1) + b\beta(\mathbf{A}_1),$$

onde

$$(6.3.1) \quad L_1(\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta_1) \quad \text{e} \quad L_0(\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta_0).$$

**Prova.** Conforme visto acima, para qualquer teste com região crítica  $\mathbf{A}_1$ , temos que

$$\alpha(\mathbf{A}_1) = \sum_{\mathbf{x} \in \mathbf{A}_1} f(\mathbf{x}|\theta_0) \quad \text{e} \quad \beta(\mathbf{A}_1) = \sum_{\mathbf{x} \in \mathbf{A}_0} f(\mathbf{x}|\theta_1),$$

para uma variável aleatória  $X$  discreta. Então,

$$\begin{aligned} a\alpha(\mathbf{A}_1) + b\beta(\mathbf{A}_1) &= a \sum_{\mathbf{x} \in \mathbf{A}_1} f(\mathbf{x}|\theta_0) + b \sum_{\mathbf{x} \in \mathbf{A}_0} f(\mathbf{x}|\theta_1) \\ &= a \sum_{\mathbf{x} \in \mathbf{A}_1} f(\mathbf{x}|\theta_0) + b \left( 1 - \sum_{\mathbf{x} \in \mathbf{A}_1} f(\mathbf{x}|\theta_1) \right) = b + \sum_{\mathbf{x} \in \mathbf{A}_1} [af(\mathbf{x}|\theta_0) - bf(\mathbf{x}|\theta_1)]. \end{aligned}$$

Portanto a soma  $a\alpha(\mathbf{A}_1) + b\beta(\mathbf{A}_1)$  será mínima quando a região crítica incluir somente os pontos amostrais  $\mathbf{x}$  tais que  $af(\mathbf{x}|\theta_0) - bf(\mathbf{x}|\theta_1) \leq 0$ , ou seja, quando

$$\frac{f(\mathbf{x}|\theta_1)}{f(\mathbf{x}|\theta_0)} = \frac{L_1(\mathbf{x})}{L_0(\mathbf{x})} \geq \frac{a}{b},$$

o que conclui a prova.

Para o caso em que  $X$  é uma variável aleatória contínua, a demonstração é análoga, bastando substituir as somas por integrais correspondentes.

**Exemplo 6.3.2.** Consideremos o Exemplo 6.3.1 novamente. Temos que o teste com  $\alpha + \beta$  ( $a = b = 1$ ) mínimo tem região crítica dada por  $\mathbf{A}_1^* = \{0, 1, 2, 3, 4\}$ , de modo que  $\alpha = 0,5$  e  $\beta = 0,3$  sendo  $\alpha + \beta = 0,80$ .

O resultado que apresentamos a seguir considera o teste mais poderoso (M.P.) de nível  $\alpha$  para testar  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta = \theta_1$ .

**Lema 6.3.2.** (Lema de Neyman-Pearson) *Consideremos o teste com região crítica*

$$(6.3.2) \quad \mathbf{A}_1^* = \left\{ \mathbf{x}; \frac{L_1(\mathbf{x})}{L_0(\mathbf{x})} \geq k \right\}.$$

em que  $L_0(\mathbf{x})$  e  $L_1(\mathbf{x})$  são dados em (6.3.1). Então  $\mathbf{A}_1^*$  é a melhor região crítica de nível  $\alpha = \alpha(\mathbf{A}_1^*)$  para testar  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta = \theta_1$ , isto é,  $\beta(\mathbf{A}_1^*) \leq \beta(\mathbf{A}_1)$  para qualquer outro teste  $\mathbf{A}_1$  com  $\alpha(\mathbf{A}_1) \leq \alpha$ .

**Prova.** Do Lema 6.3.1, temos que

$$(6.3.3) \quad k\alpha(\mathbf{A}_1^*) + \beta(\mathbf{A}_1^*) \leq k\alpha(\mathbf{A}_1) + \beta(\mathbf{A}_1),$$

para qualquer outra região crítica  $\mathbf{A}_1$ . Como  $\alpha(\mathbf{A}_1) \leq \alpha(\mathbf{A}_1^*)$ , a desigualdade (6.3.3) implica que  $\beta(\mathbf{A}_1^*) \leq \beta(\mathbf{A}_1)$ , o que conclui a prova.

O teste com região crítica (6.3.2) é também conhecido como teste da razão de verossimilhanças. Calculando a função de verossimilhança dada em (3.1.1) sob  $H_0$  ( $L_0(\mathbf{x})$ ) e sob  $H_1$  ( $L_1(\mathbf{x})$ ), o teste mais poderoso rejeita  $H_0$  quando  $L_1(\mathbf{x})/L_0(\mathbf{x}) \geq k$ , ou seja, quando a evidência em favor de  $H_1$  (expressa por  $L_1(\mathbf{x})$ ) é maior que a evidência em favor de  $H_0$  (expressa por  $L_0(\mathbf{x})$ ). Portanto, a seguir, quando nos referimos ao teste M.P., nos referimos à região crítica  $\mathbf{A}_1^*$ .

**Exemplo 6.3.3.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição de  $X \sim N(\mu, 1)$ . O objetivo é encontrar o teste M.P. para testar  $H_0 : \mu = 0$  contra  $H_1 : \mu = 1$ . Nesse caso, a função de verossimilhança é dada por

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

de modo que o teste M.P. rejeita  $H_0$  quando

$$\frac{L_1(\mathbf{x})}{L_0(\mathbf{x})} = \frac{\left( \frac{1}{\sqrt{2\pi}} \right)^n e^{-\sum_{i=1}^n (x_i - 1)^2 / 2}}{\left( \frac{1}{\sqrt{2\pi}} \right)^n e^{-\sum_{i=1}^n x_i^2 / 2}} \geq k,$$

ou seja, quando

$$e^{\sum_{i=1}^n x_i - \frac{n}{2}} \geq k,$$

que é equivalente a rejeitar  $H_0$  quando  $\sum_{i=1}^n x_i \geq \log k + n/2 = c$ . Portanto a região crítica do teste M.P. é dada por

$$(6.3.4) \quad \mathbf{A}_1^* = \left\{ \mathbf{x}; \sum_{i=1}^n x_i \geq c \right\}.$$

Dado  $\alpha = 0,05$ , por exemplo,  $c$  é tal que

$$0,05 = P_{H_0} \left[ \sum_{i=1}^n X_i \geq c \right].$$

Como, sob  $H_0$ ,  $\sum_{i=1}^n X_i \sim N(0, n)$ , temos que  $c = 1,64\sqrt{n}$ . Sendo  $n = 9$ , temos que  $c = 4,92$ , de modo que, de (6.3.4),

$$(6.3.5) \quad \mathbf{A}_1^* = \left\{ \mathbf{x}; \sum_{i=1}^n x_i \geq 4,92 \right\}.$$

Associada à região crítica (6.3.5), temos que

$$\beta = P_{H_1} \left[ \sum_{i=1}^n X_i < 4,92 \right] = P_{H_1} \left[ \frac{\sum_{i=1}^n X_i - n}{\sqrt{n}} < \frac{4,92 - n}{\sqrt{n}} \right],$$

e como  $n = 9$ ,  $\beta = P \left[ Z \leq -\frac{4,08}{3} \right] = 0,09$ , onde  $Z \sim N(0, 1)$ . O poder do teste é, então, dado por  $\pi(\theta_1) = 1 - \beta = 0,91$ . Sendo as hipóteses de interesse  $H_0 : \mu = \mu_0$  e  $H_1 : \mu = \mu_1 > \mu_0$ , o teste M.P. tem região crítica dada por (6.3.4) com  $c$  dado por

$$c = 1,64\sqrt{n} + n\mu_0.$$

**Exemplo 6.3.4.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X \sim N(\mu, \sigma^2)$ , onde  $\mu$  é conhecido. Queremos o teste M.P. para testar  $H_0 : \sigma^2 = \sigma_0^2$  contra  $H_1 : \sigma^2 = \sigma_1^2 (> \sigma_0^2)$ . De acordo com o Lema 6.3.2, temos que o teste M.P. rejeita  $H_0$  quando

$$\frac{L_1(\mathbf{x})}{L_0(\mathbf{x})} = \frac{\left( \frac{1}{\sqrt{2\pi\sigma_1^2}} \right)^n e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma_1^2}}}{\left( \frac{1}{\sqrt{2\pi\sigma_0^2}} \right)^n e^{-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma_0^2}}} \geq k,$$

que é equivalente a

$$\sum_{i=1}^n (x_i - \mu)^2 \geq \frac{\log(k(\frac{\sigma_1}{\sigma_0})^n)}{\frac{1}{2} \left( \frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} \right)} = c.$$

Então, a região crítica do teste M.P. é dada por

$$(6.3.6) \quad \mathbf{A}_1^* = \left\{ \mathbf{x}; \sum_{i=1}^n (x_i - \mu)^2 \geq c \right\}.$$

Fixando  $\alpha$ , temos que o valor de  $c$  em (6.3.6) é dado pela solução da equação

$$\alpha = P_{H_0} \left[ \sum_{i=1}^n (X_i - \mu)^2 \geq c \right] = P \left[ \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma_0^2} \geq \frac{c}{\sigma_0^2} \right].$$

Mas, sob  $H_0$ ,

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma_0^2} \sim \chi_n^2,$$

então, sendo  $\alpha = 0,05$ ,  $n = 10$  e  $\sigma_0^2 = 8$ , temos

$$0,05 = P \left[ \chi_{10}^2 \geq \frac{c}{8} \right]$$

onde  $\chi_{10}^2$  é a variável aleatória com distribuição quiquadrado com 10 graus de liberdade. Portanto temos que a região crítica é dada por

$$(6.3.7) \quad \mathbf{A}_1^* = \left\{ \mathbf{x}; \sum_{i=1}^{10} (x_i - \mu)^2 \geq 146,456 \right\}.$$

Nesse caso, sendo  $\sigma_1^2 = 10,0$  temos que

$$\beta = P_{H_1} \left[ \sum_{i=1}^{10} (X_i - \mu)^2 < 146,456 \right] = P \left[ \chi_{10}^2 \leq 14,646 \right] = 0,85,$$

pois, sob  $H_1$ ,

$$\sum_{i=1}^{10} \frac{(X_i - \mu)^2}{10} \sim \chi_{10}^2.$$

Assim, associado à região crítica (6.3.7) temos o poder  $\pi(\sigma_1^2) = 1 - \beta = 0,15$ .

**Exemplo 6.3.5.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição da variável aleatória  $X$  com distribuição *Bernoulli*( $\theta$ ). Consideremos o problema de testar  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta = \theta_1$  ( $\theta_1 > \theta_0$ ). De



acordo com o Lema de Neyman-Pearson e a função de verossimilhança dada em (3.1.1), a região crítica do teste M.P. rejeita  $H_0$  quando

$$\frac{\theta_1^{\sum_{i=1}^n x_i} (1 - \theta_1)^{n - \sum_{i=1}^n x_i}}{\theta_0^{\sum_{i=1}^n x_i} (1 - \theta_0)^{n - \sum_{i=1}^n x_i}} \geq k,$$

que pode ser escrita como

$$\left( \frac{\theta_1(1 - \theta_0)}{\theta_0(1 - \theta_1)} \right)^{\sum_{i=1}^n x_i} \geq k \left( \frac{1 - \theta_0}{1 - \theta_1} \right)^n,$$

que se reduz a

$$\sum_{i=1}^n x_i \geq \frac{\log[k(\frac{1-\theta_0}{1-\theta_1})^n]}{\log[\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}]} = c.$$

Portanto a região crítica do teste M.P. é dada por

$$\mathbf{A}_1^* = \left\{ \mathbf{x}; \sum_{i=1}^n x_i \geq c \right\}.$$

Sob  $H_0$ ,  $\sum_{i=1}^n X_i \sim \text{Binomial}(n, \theta_0)$ , então sendo  $\alpha = 0,055$ ,  $\theta_0 = 0,5$ ,  $\theta_1 = 0,6$  e  $n = 10$ , temos que

$$\alpha = P_{H_0} \left[ \sum_{i=1}^n X_i \geq c \right],$$

leva à região crítica

$$(6.3.8) \quad \mathbf{A}_1^* = \left\{ \mathbf{x}; \sum_{i=1}^{10} x_i \geq 8 \right\}.$$

Assim, associada à região crítica  $\mathbf{A}_1^*$  em (6.3.8), temos que

$$\beta = P_{H_1} \left[ \sum_{i=1}^{10} X_i \leq 7 \right] = 0,833.$$

Portanto o poder associado à região crítica (6.3.8) é dado por  $\pi(0,6) = 1 - 0,833 = 0,167$ . Sendo  $n$  grande (maior que 20, pelo menos), podemos usar a aproximação normal, ou seja,

$$\frac{\sum_{i=1}^n X_i - n\theta}{\sqrt{n\theta(1 - \theta)}} \stackrel{a}{\sim} N(0, 1).$$

Dado  $\alpha$ , podemos obter o valor de  $c$  na região crítica (6.3.8), como solução da equação

$$\alpha = P \left[ Z \geq \frac{c - n\theta_0}{\sqrt{n\theta_0(1 - \theta_0)}} \right],$$

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

Definimos a seguir nível descritivo que está associado ao valor efetivamente observado da estatística do teste.

**Definição 6.3.1.** *Consideramos como nível descritivo, que denotamos por  $\hat{\alpha}$ , como o menor nível de significância  $\alpha$  para o qual a hipótese nula  $H_0$  seria rejeitada.*

Notemos que, se  $\alpha > \hat{\alpha}$ , rejeitamos  $H_0$  e, se  $\alpha < \hat{\alpha}$ , não rejeitamos  $H_0$ , onde  $\alpha$  é o nível de significância adotado.

**Exemplo 6.3.6.** Consideremos novamente o Exemplo 6.3.3 e suponhamos que para uma amostra de  $n = 9$  observações,  $\bar{x} = 0,68$ . Portanto

$$\hat{\alpha} = P_{H_0}[\bar{X} \geq 0,68] = P[Z \geq 2,04] = 0,02,$$

onde  $Z \sim N(0, 1)$ . Nesse caso, tomando  $\alpha = 0,05$ , rejeitamos  $H_0 : \mu = 0$ .

## 6.4 Testes Uniformemente Mais Poderosos

Na seção anterior consideramos testes ótimos (M.P.) para testar hipóteses nulas simples contra alternativas simples. Nesta seção generalizamos os resultados da Seção 6.3 para o caso de hipóteses mais complexas. A Seção 6.4.1 apresenta testes ótimos para o caso em que temos hipótese nula simples e alternativas compostas. Na Seção 6.4.2, discutimos brevemente o caso em que as duas hipóteses são compostas.

### 6.4.1 Hipótese nula simples contra alternativa composta

Consideremos que as hipóteses de interesse são  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta \in \Theta_1$ .

**Definição 6.4.1.** *Um teste  $\mathbf{A}_1^*$  é dito ser uniformemente mais poderoso (U.M.P.) para testar  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta \in \Theta_1$ , se ele é M.P. de nível  $\alpha$  para testar  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta = \theta_1$ , qualquer que seja  $\theta_1 \in \Theta_1$ .*

De acordo com a Definição 6.4.1, a região crítica  $\mathbf{A}_1^*$  não pode depender particularmente de  $\theta_1$ , para qualquer  $\theta_1 \in \Theta_1$ .

**Exemplo 6.4.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição  $N(\mu, 1)$ . Consideremos as hipóteses  $H_0 : \mu = 0$  contra  $H_1 : \mu > 0$ . Neste caso,  $\Theta_1 = \{\mu; \mu > 0\}$ . Para testar  $H_0 : \mu = 0$  contra  $H_1 : \mu = \mu_1 > 0$ , temos do Exemplo 6.3.3 que o teste M.P. tem região crítica dada por  $\mathbf{A}_1^* = \{\mathbf{x}; \sum_{i=1}^n x_i \geq c\}$ . Como  $\mathbf{A}_1^*$  não depende do particular  $\mu_1$  especificado acima, segue da Definição 6.4.1 que  $\mathbf{A}_1^*$  é a região crítica do teste U.M.P. para testar  $H_0 : \mu = 0$  contra  $H_1 : \mu > 0$ .

**Exemplo 6.4.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da distribuição  $Bernoulli(\theta)$ . Consideremos as hipóteses  $H_0 : \theta = 0,5$  contra  $H_1 : \theta < 0,5$ . Para testar  $H_0 : \theta = 0,5$  contra  $H_1 : \theta = \theta_1 < 0,5$ , temos que o teste M.P. tem região crítica dada por  $\mathbf{A}_1^* = \{\mathbf{x}; \sum_{i=1}^n x_i \leq c\}$ . Como  $\mathbf{A}_1^*$  não depende do particular valor de  $\theta_1$  especificado em  $H_1$ , temos que  $\mathbf{A}_1^*$  é a região crítica do teste U.M.P. para testar  $H_0 : \theta = 0,5$  contra  $H_1 : \theta < 0,5$ .

**Exemplo 6.4.3.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\mu, 1)$ . Consideremos as hipóteses  $H_0 : \mu = 0$  contra  $H_1 : \mu \neq 0$ . Para testar  $H_0 : \mu = 0$  contra  $H_1 : \mu = 1$ , o teste M.P. é dado por  $\mathbf{A}_1^* = \{\mathbf{x}; \sum_{i=1}^n x_i \geq c\}$ . Por outro lado, para testar  $H_0 : \mu = 0$  contra  $H_1 : \mu = -1$ , o teste M.P. tem região crítica dada por  $\mathbf{A}_1^* = \{\mathbf{x}; \sum_{i=1}^n x_i \leq c\}$ . Portanto a região crítica do teste M.P. depende do particular valor de  $\mu_1$  escolhido para  $H_1$ , ou seja, a região crítica não é única. Portanto não existe teste U.M.P. para testar  $H_0 : \mu = 0$  contra  $H_1 : \mu \neq 0$ .

**Definição 6.4.2.** A função de poder  $\pi(\theta)$  com região crítica  $\mathbf{A}_1^*$  para testar  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta \in \Theta_1$  é dada por

$$\pi(\theta) = P_\theta[\mathbf{X} \in \mathbf{A}_1^*],$$

ou seja, é a probabilidade de rejeitar  $H_0$  para  $\theta \in \Theta$ . Notemos que  $\pi(\theta_0) = \alpha$ .

**Exemplo 6.4.4.** Sejam  $X_1, \dots, X_n$ , uma amostra aleatória de tamanho  $n$  da distribuição  $N(\mu, 1)$ . Consideremos o problema de testar  $H_0 : \mu = 0$  contra  $H_1 : \mu > 0$ . Conforme visto no Exemplo 6.4.1, a região crítica do teste U.M.P. é dada por  $\mathbf{A}_1^* = \{\mathbf{x}; \sum_{i=1}^n x_i \geq c\}$ . Sendo  $n = 9$  e  $\alpha = 0,05$ , temos, como no Exemplo 6.3.3, que  $c = 1,64\sqrt{9} = 4,92$ , de modo que  $\mathbf{A}_1^* = \{\mathbf{x}; \sum_{i=1}^n x_i \geq 4,92\}$ . A função de poder é, então, dada por

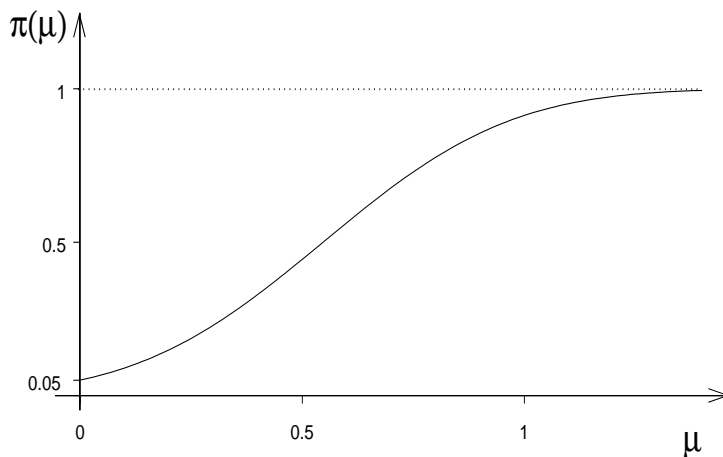
$$(6.4.1) \quad \pi(\mu) = P_\mu \left[ \sum_{i=1}^9 X_i \geq 4,92 \right] = 1 - \Phi \left( \frac{4,92 - 9\mu}{3} \right),$$

onde  $\Phi(\cdot)$  denota a função de distribuição acumulada da distribuição  $N(0, 1)$ . Então,

$$\pi(0,3) = 1 - \Phi(0,74) = 1 - 0,77 = 0,23.$$

De modo similar,  $\pi(0, 5) = 1 - \Phi(0, 14) = 0, 44$  e  $\pi(1, 0) = 0, 91$  e  $\pi(0, 0) = 0, 05 = \alpha$ . Graficamente, temos a Figura 6.1 que representa a função poder do teste.

**Figura 6.1.** Função poder dada em (6.4.1)



#### 6.4.2 Hipóteses compostas

Nesta seção consideramos brevemente testes U.M.P. para situações onde as hipóteses nula e alternativa são compostas. Mais especificamente, consideramos o problema de se testar as hipóteses  $H_0 : \theta \in \Theta_0$  contra  $H_1 : \theta \in \Theta_1$ . O resultado apresentado a seguir estabelece condições para que se tenha o teste U.M.P. para testar as hipóteses compostas acima. A demonstração pode ser vista em De Groot (1975).

**Teorema 6.4.1.** *No caso em que  $X_1, \dots, X_n$  seguem uma distribuição da família exponencial (Seção 2.4), temos que o teste U.M.P. para testar  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta > \theta_0$  é também U.M.P. para testar  $H_0 : \theta \leq \theta_0$  contra  $H_1 : \theta > \theta_0$ . Também o teste U.M.P. para testar  $H_0 : \theta = \theta_0$  contra  $H_1 : \theta < \theta_0$  é U.M.P. para testar  $H_0 : \theta \geq \theta_0$  contra  $H_1 : \theta < \theta_0$ .*

**Exemplo 6.4.5.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da variável aleatória  $X \sim N(\mu, 1)$ . De acordo com o Teorema 6.4.1, temos do Exemplo 6.4.1 que o teste U.M.P. para testar  $H_0 : \mu \leq 0$  contra  $H_1 : \mu > 0$  tem região crítica dada por  $\mathbf{A}_1^* = \{\mathbf{x}; \sum_{i=1}^n x_i \geq c\}$ .

**Exemplo 6.4.6.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim \text{Bernoulli}(\theta)$ . De acordo com o Teorema 6.4.1 e Exemplo 6.4.2, segue que

o teste U.M.P. para testar  $H_0 : \theta \geq 0,5$  contra  $H_1 : \theta < 0,5$  é dada por  $\mathbf{A}_1^* = \{\mathbf{x}; \sum_{i=1}^n x_i \leq c\}$ .

A função de poder do teste U.M.P., nesta situação mais geral, é também como na Definição 6.4.2, ou seja,  $\pi(\theta) = P_\theta[\mathbf{X} \in \mathbf{A}_1^*]$ ,  $\theta \in \Theta$ .

## 6.5 Testes da Razão de Verossimilhanças Generalizada

Na Seção 6.4 vimos que os testes UMP existem apenas em situações especiais. Essas situações compreendem o caso das famílias exponenciais unidimensionais. Vimos também que, em geral, não existem testes UMP para testar  $H_0 : \theta = \theta_0$  versus  $H_1 : \theta \neq \theta_0$ . Também não existe teste UMP na maioria dos casos em que a distribuição envolve mais de um parâmetro desconhecido como, por exemplo, a  $N(\mu, \sigma^2)$  com  $\mu$  e  $\sigma^2$  desconhecidos. Um procedimento que produz testes razoáveis e que pode ser utilizado em muitos casos, sem muita dificuldade, é o Teste da Razão de Verossimilhanças Generalizada (TRVG).

Consideremos uma situação bastante geral onde as hipóteses de interesse são

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1$$

onde  $\Theta = \Theta_0 \cup \Theta_1$ ,  $\Theta_0 \cap \Theta_1 = \emptyset$ ,  $\Theta_0 \neq \emptyset$  e  $\Theta_1 \neq \emptyset$ .

O TRVG pode ser definido como o teste com região crítica dada por (ver Bickel e Doksum(1976))

$$\mathbf{A}_1^* = \left\{ \mathbf{x}; \frac{\sup_{\theta \in \Theta_1} L(\theta; \mathbf{x})}{\sup_{\theta \in \Theta_0} L(\theta; \mathbf{x})} \geq c \right\}.$$

Podemos notar que, quando as hipóteses são simples, ou seja,  $\Theta_0 = \{\theta_0\}$  e  $\Theta_1 = \{\theta_1\}$ , o TRVG coincide com o LNP dado em (6.3.2).

Como

$$\frac{\sup_{\theta \in \Theta} L(\theta; \mathbf{x})}{\sup_{\theta \in \Theta_0} L(\theta; \mathbf{x})} = \max \left\{ 1, \frac{\sup_{\theta \in \Theta_1} L(\theta; \mathbf{x})}{\sup_{\theta \in \Theta_0} L(\theta; \mathbf{x})} \right\},$$

por facilidades computacionais o TRVG pode também ser definido como

$$(6.5.1) \quad \mathbf{A}_1^* = \left\{ \mathbf{x}; \lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta; \mathbf{x})}{\sup_{\theta \in \Theta} L(\theta; \mathbf{x})} \leq c \right\}.$$

Observemos que  $0 \leq \lambda(\mathbf{x}) \leq 1$ , pois o numerador é o supremo com relação a  $\theta$  pertencente a um subconjunto de  $\Theta$  ( $\Theta_0 \in \Theta$ ), enquanto que o denominador é o supremo sobre todo conjunto  $\Theta$ . Se a hipótese  $H_0$  for verdadeira, esperamos que  $\lambda(\mathbf{x})$  esteja “próximo” de 1, e se a hipótese  $H_0$  for falsa, esperamos que o denominador seja grande em relação ao numerador, e, portanto,  $\lambda(\mathbf{x})$  deve ser “próximo” de zero.

Para determinar  $c$  em (6.5.1) temos que resolver a equação

$$\alpha = \sup_{\theta \in \Theta_0} P(\lambda(\mathbf{X}) \leq c).$$

Para isso, precisamos da distribuição da estatística  $\lambda(\mathbf{X})$  que, em geral, não é simples de ser obtida, ou, então, podemos encontrar uma função  $h$  estritamente crescente no domínio de  $\lambda(\mathbf{x})$  tal que  $h(\lambda(\mathbf{X}))$  tenha uma forma simples e uma distribuição conhecida e tabelada sob a hipótese  $H_0$ .

Para implementação do TRVG, os seguintes passos devem ser seguidos:

- 1) obter o estimador de máxima verossimilhança (*EMV*)  $\hat{\theta}$  de  $\theta$ ;
- 2) obter o *EMV*  $\hat{\theta}_0$  de  $\theta$ , quando  $\theta \in \Theta_0$ ;
- 3) calcular  $\lambda(\mathbf{X}) = \frac{L(\hat{\theta}_0; \mathbf{X})}{L(\hat{\theta}; \mathbf{X})}$ ;
- 4) encontrar a função  $h$ ;
- 5) obter  $c$ , resolvendo a equação  $\alpha = P_{H_0}(h(\lambda(\mathbf{X})) \leq c)$ .

A seguir apresentamos alguns exemplos.

**Exemplo 6.5.1.** Consideremos o Exemplo 6.3.3 novamente, mas agora o interesse é testar  $H_0 : \mu = \mu_0$  versus  $H_1 : \mu \neq \mu_0$ . Conforme vimos no Exemplo 6.4.3 não existe teste UMP nesse caso. Pelo Exemplo 3.1.1, temos que o *EMV* de  $\mu$  é dado por  $\hat{\mu} = \bar{X}$ . Como a hipótese  $H_0$  só especifica um único valor para  $\mu$ , o numerador de  $\lambda(\mathbf{x})$  em (6.5.1) é  $L(\mu_0; \mathbf{x})$  de modo que

$$\lambda(\mathbf{x}) = \frac{(2\pi)^{-n/2} e^{-\frac{1}{2} \sum (x_i - \mu_0)^2}}{(2\pi)^{-n/2} e^{-\frac{1}{2} \sum (x_i - \bar{x})^2}} = e^{-\frac{1}{2} [\sum (x_i - \mu_0)^2 - \sum (x_i - \bar{x})^2]}.$$

Podemos simplificar  $\lambda(\mathbf{x})$  usando o fato de que

$$(6.5.2) \quad \sum (x_i - \mu_0)^2 = \sum (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2.$$

De (6.5.1) temos que o TRVG rejeita  $H_0$  quando

$$e^{-\frac{n}{2}(\mu_0 - \bar{x})^2} \leq c,$$

que é equivalente a rejeitar  $H_0$  quando

$$|\bar{x} - \mu_0| \geq \sqrt{-2 \log c / n}.$$

Portanto a região crítica do TRVG é dada por

$$\mathbf{A}_1^* = \{\mathbf{x}; \sqrt{n}|\bar{x} - \mu_0| \geq a\}.$$

Fixado  $\alpha$ , obtemos  $a$  de forma que

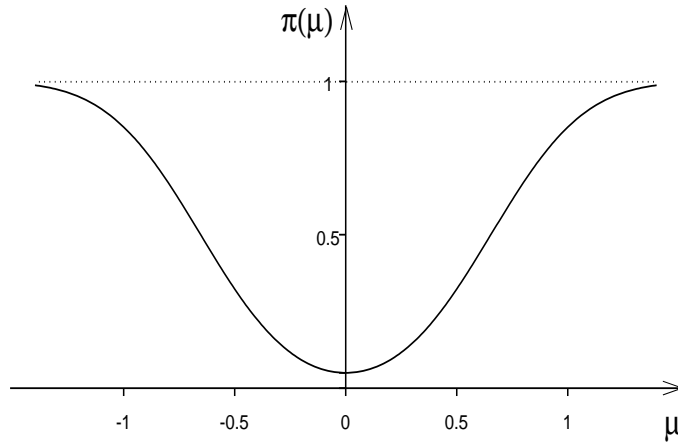
$$\alpha = P_{H_0}(\sqrt{n}|\bar{X} - \mu_0| \geq a)$$

Como sob  $H_0$ ,  $\sqrt{n}(\bar{X} - \mu_0) \sim N(0, 1)$ , temos que  $a = z_{\alpha/2}$ . Sendo  $\alpha = 0,05$  temos que  $\mathbf{A}_1^* = \{\mathbf{x}; \sqrt{n}|\bar{x} - \mu_0| \geq 1,96\}$ . Considerando  $\mu_0 = 0$ ,  $n = 9$ ,  $\sum_{i=1}^n x_i = 3,4$ , não rejeitamos  $H_0$  pois  $\sqrt{9}|3,4/9 - 0| < 1,96$ . Nesse caso, a função de poder do teste é

$$\begin{aligned} \pi(\mu) &= P_\mu(\sqrt{n}|\bar{X}| \geq 1,96) = 1 - P(-1,96 - \sqrt{n}\mu \leq \sqrt{n}(\bar{X} - \mu) \leq 1,96 - \sqrt{n}\mu) \\ &= 1 - [\Phi(1,96 - \sqrt{n}\mu) - \Phi(-1,96 - \sqrt{n}\mu)], \end{aligned}$$

pois temos que  $\sqrt{n}(\bar{X} - \mu) \sim N(0, 1)$  quando  $\mu$  é o verdadeiro valor do parâmetro. A Figura 6.2 apresenta o gráfico dessa função poder para os dados acima. Notemos que  $\pi(0) = 1 - P(-1,96 \leq Z \leq 1,96) = 0,05$ , onde  $Z \sim N(0, 1)$ . De maneira similar,  $\pi(0,3) = \pi(-0,3) = 0,15$ , e assim por diante.

**Figura 6.2.** Função poder



**Exemplo 6.5.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\mu, \sigma^2)$  com  $\mu$  e  $\sigma^2$  desconhecidos. O interesse é testar  $H_0 : \mu = \mu_0$  versus  $H_1 : \mu \neq \mu_0$ . Nesse caso,

$$\Theta_0 = \{(\mu_0, \sigma^2); \sigma^2 > 0\} \quad \text{e} \quad \Theta = \{(\mu, \sigma^2), -\infty < \mu < \infty, \sigma^2 > 0\}$$

De acordo com o Exemplo 3.4.1, o EMV de  $(\mu, \sigma^2)$  em  $\Theta$  é dado por  $\hat{\mu} = \bar{X}$  e  $\hat{\sigma}^2 = \sum (X_i - \bar{X})^2 / n$  e em  $\Theta_0$  é dado por  $\hat{\mu}_0 = \mu_0$  e  $\hat{\sigma}_0^2 = \sum (X_i - \mu_0)^2 / n$ . Logo a estatística do TRVG é dada por

$$\lambda(\mathbf{x}) = \frac{(2\pi)^{-n/2}(\hat{\sigma}_0^2)^{-n/2}e^{-\frac{1}{2\hat{\sigma}_0^2}\sum(x_i-\mu_0)^2}}{(2\pi)^{-n/2}(\hat{\sigma}^2)^{-n/2}e^{-\frac{1}{2\hat{\sigma}^2}\sum(x_i-\bar{x})^2}} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{n/2}.$$

Usando (6.5.2), temos que o TRVG rejeita  $H_0$  quando

$$\left(\frac{1}{1 + \frac{n(\bar{x}-\mu_0)^2}{\sum(x_i-\bar{x})^2}}\right)^{n/2} \leq c$$

que é equivalente a rejeitar  $H_0$  quando

$$\frac{\sqrt{n}|\bar{x}-\mu_0|}{\sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}} \geq \sqrt{(c^{-2/n}-1)(n-1)}$$

Portanto a região crítica do TRVG é dada por

$$\mathbf{A}_1^* = \left\{ \mathbf{x}; \frac{\sqrt{n}|\bar{x}-\mu_0|}{s} \geq a \right\}$$

onde  $s^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}$ . Sob a hipótese  $H_0$ ,  $\frac{\sqrt{n}(\bar{X}-\mu_0)}{S} \sim t_{n-1}$  e, então, dado  $\alpha = 0,05$  e  $n = 9$  obtemos, usando a tabela da distribuição  $t$  com 8 graus de liberdade,  $a = 2,306$ . Se  $\mu_0 = 0$ ,  $\bar{x} = 0,68$  e  $s = 1,2$ , então  $\frac{\sqrt{n}(\bar{x}-\mu_0)}{s} = 1,7$  de modo que não rejeitamos  $H_0$ .

**Exemplo 6.5.3.** Consideremos novamente o Exemplo 6.5.2, mas sendo que o interesse é testar  $H_0 : \sigma^2 = \sigma_0^2$  versus  $H_1 : \sigma^2 \neq \sigma_0^2$ . Nesse caso,

$$\Theta_0 = \{(\mu, \sigma^2); -\infty < \mu < \infty, \sigma^2 = \sigma_0^2\}$$

e

$$\Theta = \{(\mu, \sigma^2), -\infty < \mu < \infty, \sigma^2 > 0\}$$

Pelo Exemplo 3.4.1., o  $EMV$  de  $(\mu, \sigma^2)$  em  $\Theta$  é dado por  $\hat{\mu} = \bar{X}$  e  $\hat{\sigma}^2 = \sum(X_i - \bar{X})^2/n$ , enquanto que em  $\Theta_0$  é dado por  $\hat{\mu}_0 = \bar{X}$  e  $\hat{\sigma}_0^2 = \sigma_0^2$ . Logo, a estatística do TRVG é dada por

$$\lambda(\mathbf{x}) = \frac{(2\pi)^{-n/2}(\sigma_0^2)^{-n/2}e^{-\frac{1}{2\sigma_0^2}\sum(x_i-\bar{x})^2}}{(2\pi)^{-n/2}(\hat{\sigma}^2)^{-n/2}e^{-\frac{1}{2\hat{\sigma}^2}\sum(x_i-\bar{x})^2}} = \left(\frac{\hat{\sigma}^2}{\sigma_0^2}\right)^{n/2} e^{-\frac{1}{2\sigma_0^2}\sum(x_i-\bar{x})^2 + n/2}.$$

Então, temos que o TRVG rejeita  $H_0$  quando

$$\left(\frac{\sum(x_i-\bar{x})^2}{\sigma_0^2}\right)^{n/2} e^{-\frac{\sum(x_i-\bar{x})^2}{2\sigma_0^2}} \leq c.$$



Notemos que se  $g(y) = y^{n/2}e^{-y/2}$ ,  $y > 0$  então a função  $\log g(y)$  (e também  $g(y)$ ) é crescente para  $y < n$ , atingindo o ponto de máximo em  $y = n$  e é decrescente para  $y > n$ , logo  $g(y) \leq c$  se e somente se  $y \leq c_1$  ou  $y \geq c_2$  com  $g(c_1) = g(c_2)$ . Portanto o TRVG é equivalente a rejeitar  $H_0$  quando

$$\frac{\sum (x_i - \bar{x})^2}{\sigma_0^2} \leq c_1 \quad \text{ou} \quad \frac{\sum (x_i - \bar{x})^2}{\sigma_0^2} \geq c_2.$$

Sob a hipótese  $H_0$ ,  $\frac{\sum (X_i - \bar{X})^2}{\sigma_0^2} \sim \chi_{n-1}^2$  e, então, dado  $\alpha = 0,05$  e  $n = 9$  obtemos, usando a tabela da distribuição quiquadrado com 8 graus de liberdade,  $c_1 = 2,180$  e  $c_2 = 17,534$  se considerarmos, como na Seção 5.2, probabilidades iguais para as duas caudas.

**Exemplo 6.5.4.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com função densidade de probabilidade dada por

$$f(x|\theta) = \begin{cases} e^{-(x-\theta)}, & x \geq \theta \\ 0, & x < \theta \end{cases}$$

onde  $-\infty < \theta < \infty$ . A função de verossimilhança pode ser escrita como

$$L(\theta; \mathbf{x}) = \begin{cases} e^{-\sum x_i + n\theta}, & \theta \leq x_{(1)} \\ 0, & \theta > x_{(1)} \end{cases}.$$

Suponhamos que o interesse seja testar  $H_0 : \theta \leq \theta_0$  versus  $H_1 : \theta > \theta_0$  onde  $\theta_0$  é um valor especificado. Podemos verificar que  $L(\theta; \mathbf{x})$  é uma função crescente em  $\theta$  no intervalo  $-\infty < \theta \leq x_{(1)}$ . Logo, em  $\Theta$ , o EMV de  $\theta$  é  $\hat{\theta} = X_{(1)}$  e em  $\Theta_0$  é dado por  $\hat{\theta} = \theta_0$  se  $x_{(1)} > \theta_0$  e  $\hat{\theta} = x_{(1)}$  se  $x_{(1)} \leq \theta_0$ . Portanto a estatística do TRVG é dada por

$$\lambda(\mathbf{x}) = \begin{cases} 1, & x_{(1)} \leq \theta_0 \\ e^{-n(x_{(1)} - \theta_0)}, & x_{(1)} > \theta_0 \end{cases}.$$

Portanto a região crítica do TRVG pode ser escrita como

$$A_1 = \left\{ \mathbf{x}, x_{(1)} \geq \theta_0 - \frac{\log c}{n} \right\}.$$

Como mencionado anteriormente, a forma e a distribuição de  $\lambda(\mathbf{X})$  podem ser complicadas e nem sempre podemos encontrar uma função  $h$  com distribuição conhecida. O Teorema a seguir fornece a distribuição assintótica da estatística do TRVG, resolvendo esse problema pelo menos para o caso de amostras grandes. A prova desse resultado envolve conhecimentos avançados de probabilidade e pode ser encontrada em Sen e Singer (1993).

**Teorema 6.5.1.** *Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com f.d.p.  $f(x|\theta)$ . Sob as condições de regularidade, se  $\theta \in \Theta_0$ , então a distribuição da estatística  $-2\log\lambda(\mathbf{X})$  converge para a distribuição quiquadrado quando o tamanho da amostra  $n$  tende ao infinito. O número de graus de liberdade da distribuição limite é a diferença entre o número de parâmetros não especificados em  $\Theta$  e o número de parâmetros não especificados em  $\Theta_0$ .*

**Exemplo 6.5.5.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim \text{Poisson}(\theta)$ . O interesse é testar  $H_0 : \theta = 5$  versus  $H_1 : \theta \neq 5$ . Pelo Exemplo 3.2.5 temos que o EMV de  $\theta$  é dado por  $\hat{\theta} = \bar{X}$ . Como a hipótese  $H_0$  só especifica um único valor para  $\theta$ , o numerador de  $\lambda(\mathbf{x})$  em 6.5.1 é  $L(5, \mathbf{x})$  de modo que

$$\lambda(\mathbf{x}) = \frac{e^{-5n} 5^{\sum x_i}}{\prod x_i!} \frac{\prod x_i!}{e^{-n\bar{x}} \bar{x}^{\sum x_i}} = e^{-n(5-\bar{x})} (5/\bar{x})^{\sum x_i}$$

Pelo Teorema 6.5.1 temos que

$$-2\log\lambda(\mathbf{x}) = -2 \left\{ -n(5 - \bar{x}) + \sum x_i \log(5/\bar{x}) \right\}.$$

Portanto a região crítica do TRVG é dada por

$$\mathbf{A}_1^* = \{-2[-n(5 - \bar{x}) + \sum x_i \log 5/\bar{x}] \geq c\}$$

onde um valor aproximado para  $c$  é obtido de modo que  $P(\chi_1^2 \geq c) = 0,05$ , que requer a utilização da tabela da distribuição quiquadrado.

A seguir apresentamos alguns exemplos onde o interesse é a comparação de duas populações.

**Exemplo 6.5.6.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\mu_X, \sigma^2)$  e  $Y_1, \dots, Y_m$  uma amostra aleatória da variável aleatória  $Y \sim N(\mu_Y, \sigma^2)$ . Suponhamos que as amostras são independentes e que o interesse é testar  $H_0 : \mu_X = \mu_Y$  versus  $H_1 : \mu_X \neq \mu_Y$ . Nesse caso

$$\Theta_0 = \{(\mu_X, \mu_Y, \sigma^2); \mu_X = \mu_Y = \mu, -\infty < \mu < \infty, \sigma^2 > 0\}$$

e

$$\Theta = \{(\mu_X, \mu_Y, \sigma^2), -\infty < \mu_X < \infty, -\infty < \mu_Y < \infty, \sigma^2 > 0\}$$

Em  $\Theta$  os EMVs são dados por

$$\hat{\mu}_X = \bar{X} \quad , \quad \hat{\mu}_Y = \bar{Y}$$

e

$$\hat{\sigma}^2 = \frac{\sum (X_i - \bar{X})^2 + \sum (Y_i - \bar{Y})^2}{n + m},$$

enquanto que em  $\Theta_0$  são dados por

$$\hat{\mu}_0 = \frac{\sum X_i + \sum Y_i}{n + m} \quad \text{e} \quad \hat{\sigma}_0^2 = \frac{\sum (X_i - \hat{\mu}_0)^2 + \sum (y_i - \hat{\mu}_0)^2}{n + m}.$$

Logo a estatística do TRVG pode ser escrita como

$$\begin{aligned} \lambda(\mathbf{x}, \mathbf{y}) &= \frac{(2\pi)^{-(n+m)/2} (\hat{\sigma}_0^2)^{-(n+m)/2} e^{-\frac{1}{2\hat{\sigma}_0^2} \{ \sum (x_i - \hat{\mu}_0)^2 + \sum (y_i - \hat{\mu}_0)^2 \}}}{(2\pi)^{-(n+m)/2} (\hat{\sigma}^2)^{-(n+m)/2} e^{-\frac{1}{2\hat{\sigma}^2} \{ \sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2 \}}} \\ &= \left( \frac{\hat{\sigma}^2}{\hat{\sigma}_0^2} \right)^{(n+m)/2}. \end{aligned}$$

Usando (6.5.1), temos que o TRVG rejeita  $H_0$  quando

$$\left( \frac{1}{1 + \frac{n(\bar{x} - \hat{\mu}_0)^2 + m(\bar{y} - \hat{\mu}_0)^2}{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}} \right)^{(n+m)/2} \leq c$$

que é equivalente a rejeitar  $H_0$  quando

$$\frac{n(\bar{x} - \hat{\mu}_0)^2 + m(\bar{y} - \hat{\mu}_0)^2}{s_p^2} \geq c_1$$

onde  $s_p^2 = \frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{n+m-2}$ . Mas

$$\bar{x} - \hat{\mu}_0 = \frac{m}{n+m} (\bar{x} - \bar{y})$$

$$\bar{y} - \hat{\mu}_0 = \frac{n}{n+m} (\bar{y} - \bar{x}),$$

portanto a região crítica do TRVG é dada por

$$\mathbf{A}_1^* = \left\{ (\mathbf{x}, \mathbf{y}); \frac{\bar{x} - \bar{y}}{s_p \sqrt{(\frac{1}{n} + \frac{1}{m})}} \leq c_1 \quad \text{ou} \quad \frac{\bar{x} - \bar{y}}{s_p \sqrt{(\frac{1}{n} + \frac{1}{m})}} \geq c_2 \right\}$$

Sob a hipótese  $H_0$ ,  $\frac{\bar{X} - \bar{Y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}$ . Os valores de  $c_1$  e  $c_2$  são obtidos utilizando a tabela da distribuição  $t$  com  $n + m - 2$  graus de liberdade.

**Exemplo 6.5.7.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\mu_X, \sigma_X^2)$  e  $Y_1, \dots, Y_m$  uma amostra aleatória da variável aleatória  $Y \sim$

$N(\mu_Y, \sigma_Y^2)$ . Suponhamos que as amostras são independentes e que o interesse é testar  $H_0 : \sigma_X^2 = \sigma_Y^2$  versus  $H_1 : \sigma_X^2 \neq \sigma_Y^2$ . Nesse caso

$$\Theta_0 = \{(\mu_X, \mu_Y, \sigma^2); -\infty < \mu_X, \mu_Y < \infty, \sigma^2 > 0\}$$

e

$$\Theta = \{(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2), -\infty < \mu_X, \mu_Y < \infty, \sigma_X^2 > 0, \sigma_Y^2 > 0\}$$

Em  $\Theta$  os *EMVs* dos parâmetros são dados por

$$\hat{\mu}_X = \bar{X} \quad , \quad \hat{\mu}_Y = \bar{Y}$$

e

$$\hat{\sigma}_X^2 = \frac{\sum (X_i - \bar{X})^2}{n}, \quad \hat{\sigma}_Y^2 = \frac{\sum (Y_i - \bar{Y})^2}{m}$$

enquanto que em  $\Theta_0$  são dados por

$$\hat{\mu}_X = \bar{X}, \quad \hat{\mu}_Y = \bar{Y}, \quad \hat{\sigma}^2 = \frac{\sum (X_i - \bar{X})^2 + \sum (Y_i - \bar{Y})^2}{n + m}.$$

Logo a estatística do TRVG é

$$\begin{aligned} \lambda(\mathbf{x}, \mathbf{y}) &= \frac{(2\pi)^{-(n+m)/2} (\hat{\sigma}^2)^{-(n+m)/2} e^{-\frac{1}{2\hat{\sigma}^2} \{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2\}}}{(2\pi \hat{\sigma}_X^2)^{-n/2} e^{-\frac{1}{2\hat{\sigma}_X^2} \sum (x_i - \bar{x})^2} (2\pi \hat{\sigma}_Y^2)^{-m/2} e^{-\frac{1}{2\hat{\sigma}_Y^2} \sum (y_i - \bar{y})^2}} \\ &= \frac{(\hat{\sigma}_X^2)^{n/2} (\hat{\sigma}_Y^2)^{m/2}}{(\hat{\sigma}^2)^{(n+m)/2}}, \end{aligned}$$

de modo que rejeitamos  $H_0$  quando

$$g(F) = \frac{\left(\frac{m-1}{n-1} F\right)^{m/2}}{\left(1 + \frac{m-1}{n-1} F\right)^{n+m/2}} \leq c$$

onde  $F = \frac{\sum (y_i - \bar{y})^2 / (m-1)}{\sum (x_i - \bar{x})^2 / (n-1)}$ . Mas  $g(F) \leq c$  se e somente se  $F \leq c_1$  ou  $F \geq c_2$ , portanto a região crítica do TRVG é dada por

$$\mathbf{A}_1^* = \{(\mathbf{x}, \mathbf{y}); F \leq c_1 \quad \text{ou} \quad F \geq c_2\}$$

Sob a hipótese  $H_0$ ,  $F \sim F_{m-1, n-1}$  e, então, dado  $\alpha = 0,10$ ,  $m = 9$  e  $n = 8$ , obtemos usando a tabela da distribuição  $F$  com 8 e 7 graus de liberdade que  $c_1 = 0,27$  e  $c_2 = 3,5$ .

**Exemplo 6.5.8.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim \text{Bernoulli}(\theta_1)$  e  $Y_1, \dots, Y_m$  uma amostra aleatória da variável aleatória

$Y \sim \text{Bernoulli}(\theta_2)$ . Suponhamos que as amostras são independentes e que o interesse é testar  $H_0 : \theta_1 = \theta_2$  versus  $H_1 : \theta_1 \neq \theta_2$ . Nesse caso

$$\Theta_0 = \{(\theta_1, \theta_2); \theta_1 = \theta_2 = \theta, 0 < \theta < 1\}$$

e

$$\Theta = \{(\theta_1, \theta_2); 0 < \theta_1 < 1, 0 < \theta_2 < 1\}$$

Em  $\Theta$  os *EMVs* são dados por

$$\hat{\theta}_1 = \bar{X} \quad \text{e} \quad \hat{\theta}_2 = \bar{Y},$$

enquanto que em  $\Theta_0$  é dado por

$$\hat{\theta} = \frac{\sum x_i + \sum y_i}{n + m}.$$

Logo

$$\lambda(\mathbf{x}, \mathbf{y}) = \frac{\hat{\theta}^{\sum x_i + \sum y_i} (1 - \hat{\theta})^{(n+m - \sum x_i - \sum y_i)}}{\hat{\theta}_1^{\sum x_i} (1 - \hat{\theta}_1)^{n - \sum x_i} \hat{\theta}_2^{\sum y_i} (1 - \hat{\theta}_2)^{m - \sum y_i}}$$

Como não conseguimos explicitar a região crítica através de uma estatística com distribuição conhecida, então pelo Teorema 6.5.1, temos que

$$\begin{aligned} -2 \log \lambda(\mathbf{x}, \mathbf{y}) = & -2 \left( \left( \sum x_i + \sum y_i \right) \log \hat{\theta} \right. \\ & + (m + n - \sum x_i - \sum y_i) \log(1 - \hat{\theta}) \\ & - \sum x_i \log \hat{\theta}_1 - (n - \sum x_i) \log(1 - \hat{\theta}_1) \\ & \left. - \sum y_i \log \hat{\theta}_2 - (m - \sum y_i) \log(1 - \hat{\theta}_2) \right) \end{aligned}$$

tem distribuição aproximadamente  $\chi_1^2$ . Logo, quando  $-2 \log \lambda(\mathbf{x}, \mathbf{y}) \geq c$  rejeitamos  $H_0$ . Suponhamos que  $n = 400$ ,  $\sum x_i = 60$ ,  $m = 225$ ,  $\sum y_i = 40$ . Assim,  $\hat{\theta} = 100/625$  de modo que  $-2 \log \lambda(\mathbf{x}, \mathbf{y}) = 0,82$ . Tomando  $\alpha = 0,05$ , temos que  $c = 3,841$ , portanto não rejeitamos  $H_0$ .

**Exemplo 6.5.9.** Consideramos neste exemplo uma extensão do modelo binomial considerado no exemplo anterior. Suponhamos que os indivíduos em uma população podem ser de três tipos, que rotulamos por tipos 1, 2 e 3. No caso de preferência eleitoral, por exemplo, um indivíduo é do tipo 1 se ele for eleitor do partido A; do tipo 2 se for eleitor do partido B e do tipo 3 se for eleitor de um outro partido, que não o A e ou o B. Suponhamos que a proporção de indivíduos do tipo  $i$  seja  $\theta_i$ ,  $i = 1, 2, 3$ , de modo que  $\theta_1 + \theta_2 + \theta_3 = 1$ . Para uma amostra de  $n$  indivíduos observados na população suponhamos que  $n_i$  seja do

tipo  $i$ ,  $i = 1, 2, 3$ , de modo que  $n_1 + n_2 + n_3 = n$ . A função de verossimilhança pode então ser escrita como

$$(6.5.4) \quad L(\theta, \mathbf{x}) = \theta_1^{n_1} \theta_2^{n_2} (1 - \theta_1 - \theta_2)^{n - n_1 - n_2},$$

onde  $\mathbf{x} = (x_1, \dots, x_n)$ , com  $x_i$  representando o rótulo (1, 2 ou 3) do  $i$ -ésimo indivíduo observado na amostra. Portanto, como no Exemplo 3.5.1,  $n_1$ ,  $n_2$  e  $n_3$  representam o número de elementos de  $\{x_1, \dots, x_n\}$  iguais a 1, 2 ou 3, respectivamente. Derivando-se o logaritmo da verossimilhança (6.5.4) com relação a  $\theta_1$  e a  $\theta_2$ , temos os estimadores de máxima verossimilhança

$$(6.5.5) \quad \hat{\theta}_1 = \frac{n_1}{n} \quad \text{e} \quad \hat{\theta}_2 = \frac{n_2}{n},$$

de modo que o estimador de máxima verossimilhança de  $\theta_3$  é dado por  $\hat{\theta}_3 = n_3/n$  (veja o Exercício 6.13). A extensão para o caso geral (caso multinomial, com  $k$  tipos diferentes de indivíduos) pode ser feita de maneira similar. Suponhamos agora que queremos testar a hipótese de que os indivíduos na população seguem o equilíbrio de Hardy-Weinberg, isto é, que  $H_0 : \theta_1 = p(1; \theta) = \theta^2$ ,  $\theta_2 = p(2; \theta) = 2\theta(1 - \theta)$ ,  $\theta_3 = p(3; \theta) = (1 - \theta)^2$ , para  $0 < \theta < 1$ . Sob o modelo geral, ou seja, em  $\Theta = \{(\theta_1, \theta_2, \theta_3); \theta_i > 0, \theta_1 + \theta_2 + \theta_3 = 1\}$  os estimadores de máxima verossimilhança de  $\theta = (\theta_1, \theta_2, \theta_3)$  são como dados em (6.5.5). Sob a hipótese  $H_0$ , ou seja em  $\Theta_0$  (escreva!), temos que o estimador de máxima verossimilhança de  $\theta$  é obtido no Exemplo 3.5.1, ou seja, é dado por  $\hat{\theta} = (2n_1 + n_2)/2n$ . Temos, portanto, que a razão de verossimilhanças generalizada é dada por

$$\lambda(\mathbf{x}) = \frac{\left(\frac{2n_1+n_2}{2n}\right)^{2n_1} \left(2\frac{(2n_1+n_2)}{2n}\left(1 - \frac{2n_1+n_2}{2n}\right)\right)^{n_2} \left(1 - \frac{2n_1+n_2}{2n}\right)^{2n_3}}{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_2}{n}\right)^{n_2} \left(\frac{n_3}{n}\right)^{n_3}},$$

de modo que

$$(6.5.6) \quad -2 \log \lambda(\mathbf{x}) = -2 \left\{ (2n_1 + n_2) \log \left( \frac{2n_1 + n_2}{2n} \right) - n_1 \log n_1 - n_2 \log n_2 \right. \\ \left. + (n_2 + 2n_3) \log \left( 1 - \frac{2n_1 + n_2}{2n} \right) - n_3 \log n_3 + n \log n + n_2 \log 2 \right\},$$

que tem, aproximadamente, distribuição  $\chi_1^2$ .

Uma estatística assintoticamente (em grandes amostras) equivalente (veja Bickel e Doksun, 1977) à estatística da razão de verossimilhanças generalizada, calculada acima, é dada pela estatística qui-quadrado de Pearson, que no caso do modelo do equilíbrio de Hardy-Weinberg, é dada por

$$\begin{aligned}
(6.5.7) \quad Q &= \sum_{i=1}^3 \frac{(n_i - np(i; \hat{\theta}))^2}{np(i; \hat{\theta})} \\
&= \frac{(n_1 - n\hat{\theta}^2)^2}{n\hat{\theta}^2} + \frac{(n_2 - n2\hat{\theta}(1 - \hat{\theta}))^2}{n2\hat{\theta}(1 - \hat{\theta})} + \frac{(n_3 - n(1 - \hat{\theta})^2)^2}{n(1 - \hat{\theta})^2},
\end{aligned}$$

que, para  $n$  grande, tem a mesma distribuição que  $-2 \log \lambda(\mathbf{x})$ , ou seja,  $\chi_1^2$ . Notemos que a estatística  $Q$  dada em (6.5.7) é, em geral, interpretada como a soma do quadrado da diferença entre o número observado (dado por  $n_i$ ) e o número esperado (sob  $H_0$ ) de indivíduos do tipo  $i$  na amostra, que é dado por  $ng_i(\hat{\theta})$ , dividido pelo número esperado (sob  $H_0$ ) de indivíduos do tipo  $i$  na amostra, para todos os tipos de indivíduos na população. No caso do equilíbrio de Hardy-Weinberg, temos que  $p(1; \theta) = \theta^2$ ,  $p(2; \theta) = 2\theta(1 - \theta)$  e  $p(3; \theta) = (1 - \theta)^2$ . A estatística  $Q$  pode também ser generalizada para situações mais complexas que aquela considerada acima. Entre outras, citamos sua utilização em testes de independência em tabelas de contingência, discutido em textos básicos de estatística como, por exemplo, em Bussab e Morettin (1987).

Vamos discutir brevemente as relações entre testes de hipóteses e intervalos de confiança. Consideremos o Exemplo 6.5.1 novamente. Nesse exemplo temos que, para um nível  $\alpha$  fixado, a hipótese  $H_0$  é aceita se  $|\bar{x} - \mu_0| \leq z_{\alpha/2}/\sqrt{n}$ , ou equivalentemente, se

$$\bar{x} - \frac{z_{\alpha/2}}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + \frac{z_{\alpha/2}}{\sqrt{n}}.$$

Como o teste tem nível  $\alpha$ , a  $P(H_0 \text{ ser aceita} | \mu = \mu_0) = 1 - \alpha$ , então podemos escrever que

$$P\left(\bar{X} - \frac{z_{\alpha/2}}{\sqrt{n}} \leq \mu_0 \leq \bar{X} + \frac{z_{\alpha/2}}{\sqrt{n}} | \mu = \mu_0\right) = 1 - \alpha.$$

No entanto essa probabilidade deve valer para todo  $\mu_0$ , de modo que

$$P\left(\bar{X} - \frac{z_{\alpha/2}}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{z_{\alpha/2}}{\sqrt{n}}\right) = 1 - \alpha.$$

Portanto o intervalo  $\left[\bar{x} - \frac{z_{\alpha/2}}{\sqrt{n}}; \bar{x} + \frac{z_{\alpha/2}}{\sqrt{n}}\right]$  obtido a partir da região de aceitação do teste de nível  $\alpha$ , é um intervalo de  $100(1 - \alpha)\%$  de confiança para  $\mu$  e coincide com o intervalo (5.3.2).

Por outro lado, a partir do intervalo de confiança, podemos construir um teste bilateral ( $H_0 : \theta = \theta_0$  versus  $H_1 : \theta \neq \theta_0$ ) onde

$$\text{rejeitamos } H_0 \text{ se } \theta_0 \notin \text{I.C.}$$

aceitamos  $H_0$  se  $\theta_0 \in I.C.$

Esse teste tem nível  $\alpha$ , pois

$$P(H_0 \text{ ser rejeitada} | \theta = \theta_0) = P_{\theta_0}(\theta_0 \notin I.C.) = \alpha.$$

Concluimos, então, que podemos obter um intervalo de confiança a partir de um teste de hipótese e vice e versa.

## 6.6 Testes Bayesianos

O problema de testes de hipóteses também pode ser formulado do ponto de vista Bayesiano. Nesse caso, o teste será baseado na distribuição a posteriori. Como vimos na seção anterior existe uma relação entre testes de hipóteses e intervalos de confiança, então uma maneira de se construir um teste Bayesiano é através da obtenção de um intervalo de confiança Bayesiano.

Suponhamos que o interesse seja testar  $H_0 : \theta = \theta_0$  versus  $H_1 : \theta \neq \theta_0$ . Para isso, construímos o intervalo Bayesiano para  $\theta$  e, se  $\theta_0$  estiver contido no intervalo, então aceitamos  $H_0$  e, se  $\theta_0$  estiver fora do intervalo, então rejeitamos  $H_0$ .

**Exemplo 6.6.1.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\mu, 1)$ , e consideremos uma priori  $N(0, 1)$ . O interesse é testar  $H_0 : \mu = 0$  versus  $H_1 : \mu \neq 0$ . Do Exemplo 4.4.3 temos que a distribuição a posteriori de  $\mu$  é  $N\left(\frac{n\bar{x}}{n+1}, \frac{1}{n+1}\right)$ , ou seja,

$$\frac{\mu - \frac{n\bar{x}}{n+1}}{\sqrt{\frac{1}{n+1}}} \sim N(0, 1).$$

Logo

$$P\left(-z_{\alpha/2} \leq \frac{\mu - \frac{n\bar{x}}{n+1}}{\sqrt{\frac{1}{n+1}}} \leq z_{\alpha/2}\right) = \gamma$$

de modo que o intervalo Bayesiano (intervalo de credibilidade) com probabilidade  $\gamma$  é dado por

$$\left[ \frac{n\bar{x}}{n+1} - z_{\alpha/2} \sqrt{\frac{1}{n+1}}, \frac{n\bar{x}}{n+1} + z_{\alpha/2} \sqrt{\frac{1}{n+1}} \right].$$

Suponhamos que  $n = 8$ ,  $\sum_{i=1}^8 x_i = 0,57$  e  $\alpha = 0,05$ . Logo o intervalo de confiança Bayesiano é  $[-0,59; 0,72]$ . Como o zero está contido no intervalo, não rejeitamos a hipótese  $H_0$ , ao nível de  $\alpha = 5\%$ .



## 6.7 Exercícios

**6.1.** Seja  $X$  uma variável aleatória com função de densidade  $f(x|\theta) = \theta^2 x e^{-\theta x}$ ,  $x > 0, \theta > 0$ . Queremos testar  $H_0 : \theta = 1$  versus  $H_1 : \theta = 2$ .

- i) Qual é a região crítica se  $n = 5$  e  $\alpha = 0,05$ ?
- ii) Se  $n = 1$ , qual é o teste que minimiza  $\alpha + \beta$ ? E qual o valor de  $\alpha + \beta$ ?

**6.2.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\mu, 1)$ . Queremos testar  $H_0 : \mu = 0$  versus  $H_1 : \mu = 1$ . Encontre  $n$  que produz o teste mais poderoso com  $\alpha = \beta = 0,05$ .

**6.3.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com função de densidade dada por

$$f(x|\theta) = \theta x^{\theta-1}, \quad 0 < x < 1, \quad \theta > 0.$$

- i) Mostre que o teste mais poderoso para testar  $H_0 : \theta = 1$  versus  $H_1 : \theta = 2$ , rejeita  $H_0$ , se e somente se,  $\sum_{i=1}^n -\log x_i \leq a$ , onde  $a$  é uma constante.
- ii) Sendo  $n = 2$  e  $\alpha = (1 - \log 2)/2$ , qual a região crítica?

**6.4.** Seja  $X$  uma única observação da função de densidade

$$f(x|\theta) = (2\theta x + 1 - \theta)I_{(0,1)}(x)$$

Queremos testar  $H_0 : \theta = 0$  versus  $H_1 : \theta = 1$ .

- i) Obtenha o teste mais poderoso com nível de significância  $\alpha$ .
- ii) Se  $\alpha = 0,05$  e  $x = 0,8$ , qual a sua conclusão?

**6.5.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim \text{Poisson}(\theta)$ .

- i) Encontre o teste UMP para testar  $H_0 : \theta = \theta_0$  versus  $H_1 : \theta > \theta_0$ .
- ii) Seja  $\alpha = 0,05$ , faça o gráfico da função poder para  $\theta_0 = 1$  e  $n = 25$  (use o Teorema do limite central).

**6.6.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\mu_X, 1)$  e sejam  $Y_1, \dots, Y_m$  uma amostra aleatória da variável aleatória  $Y \sim N(\mu_Y, 4)$  sendo as amostras independentes.

- i) Determine o teste mais poderoso para testar

$$H_0 : \mu_X = \mu_Y = 0 \quad \text{versus} \quad H_1 : \mu_X = \mu_Y = 1$$

- ii) Sendo  $n = 9$ ,  $\sum x_i = 3,95$ ;  $m = 4$ ;  $\sum y_i = 2,03$ . Qual a sua conclusão ao nível de significância de 5%? E qual o poder do teste?

**6.7.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X$  com f.d.p. dada por

$$f(x|\theta) = \frac{1}{\theta} x^{(1-\theta)/\theta}, \quad 0 < x < 1, \quad \theta > 0.$$

Queremos testar  $H_0 : \theta \leq \theta_0$  versus  $H_1 : \theta > \theta_0$ .

- i) Encontre o teste UMP de nível  $\alpha$  (se existir).
- ii) Se  $n = 2$ ,  $\theta_0 = 1$  e  $\alpha = 0,05$ , encontre a região crítica.

**6.8.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(0, \sigma^2)$ .

- i) Encontre o teste UMP para testar  $H_0 : \sigma^2 = \sigma_0^2$  versus  $H_1 : \sigma^2 > \sigma_0^2$ .
- ii) Seja  $\alpha = 0,05$ ,  $n = 9$  e  $\sigma_0^2 = 9$ , faça o gráfico da função poder.

**6.9.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim \exp(\theta)$ .

- i) Encontre o teste da razão de verossimilhanças generalizada para testar

$$H_0 : \theta = 1 \quad \text{versus} \quad H_1 : \theta \neq 1.$$

- ii) Se você observar  $n = 5$ ;  $x_1 = 0,8$ ;  $x_2 = 1,3$ ;  $x_3 = 1,8$ ;  $x_4 = 0,9$  e  $x_5 = 1,0$ , qual a sua decisão ao nível de 5%?

**6.10.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim N(\mu_X, 9)$  e seja  $Y_1, \dots, Y_m$  uma amostra aleatória da variável aleatória  $Y \sim N(\mu_Y, 25)$ , sendo as amostras independentes.

- i) Determine o teste da RVG para testar

$$H_0 : \mu_X = \mu_Y \quad \text{versus} \quad H_1 : \mu_X \neq \mu_Y$$

- ii) Sendo  $n = 9$ ,  $\sum x_i = 3,4$ ,  $m = 16$ ,  $\sum y_i = 4,3$ . Qual a sua conclusão a um nível de significância de 5%?

**6.11.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim \text{Poisson}(\theta_1)$  e sejam  $Y_1, \dots, Y_m$  uma amostra aleatória da variável aleatória  $Y \sim \text{Poisson}(\theta_2)$  sendo as amostras independentes.

- i) Encontre o teste da RVG (aproximado) para testar  $H_0 : \theta_1 = \theta_2$  versus  $H_1 : \theta_1 \neq \theta_2$ .
- ii) Sendo  $n = 5$ ,  $\sum x_i = 3,8$ ;  $m = 8$ ;  $\sum y_i = 4,8$ , qual a sua conclusão a um nível de significância de 5%?

**6.12.** Sejam  $X_1, \dots, X_n$  uma amostra aleatória da variável aleatória  $X \sim \exp(\theta_1)$  e sejam  $Y_1, \dots, Y_n$  uma amostra aleatória da variável aleatória  $Y \sim \exp(\theta_2)$ , sendo as amostras independentes.

- i) Determine o teste mais poderoso para testar

$$H_0 : \theta_1 = \theta_2 = 1 \quad \text{versus} \quad H_1 : \theta_1 = \theta_2 = 2.$$

- ii) Verifique se seu teste é UMP para testar

$$H_0 : \theta_1 = \theta_2 = 1 \quad \text{versus} \quad H_1 : \theta_1 = \theta_2 > 1.$$

iii) Se você observar  $n = 5$ ,  $\bar{x} = 1, 1$ ;  $\bar{y} = 0, 8$ , qual a sua decisão ao nível de 5%?

iv) Determine o teste da RVG para testar  $H_0 : \theta_1 = \theta_2$  versus  $H_1 : \theta_1 \neq \theta_2$ .

v) Mostre que o teste acima é equivalente a um teste F exato.

**6.13.** Discuta a obtenção dos estimadores de máxima verossimilhança dados em (6.5.5). Suponha que em uma população com três tipos de indivíduos, temos para uma amostra de  $n = 100$  indivíduos,  $n_1 = 26$  do tipo 1,  $n_2 = 47$  do tipo 2 e  $n_3 = 27$  do tipo 3. Verifique ao nível de 5% se a distribuição dos tipos de indivíduos na população segue o equilíbrio de Hardy-Weinberg.

**6.14.** Discuta a implementação de um procedimento (teste) para verificar se um dado é equilibrado, ou seja, para testar  $H_0 : \theta_1 = \dots = \theta_6$  sendo que  $n$  lançamentos do dado apresenta  $n_i$  ocorrência da face  $i$ ,  $i = 1, \dots, 6$ . Sendo  $n = 120$ ,  $n_1 = 23$ ,  $n_2 = 18$ ,  $n_3 = 15$ ,  $n_4 = 21$ ,  $n_5 = 27$  e  $n_6 = 16$ , qual sua decisão ao nível de 5%?

**6.15.** Um modelo genético para a distribuição dos tipos de sangue 1, 2, 3 e 4, especifica as proporções  $\theta_1 = p(1; \theta) = (2 + \theta)/4$ ,  $\theta_2 = p(2; \theta) = (1 - \theta)/4 = \theta_3 = p(3; \theta)$  e  $\theta_4 = p(4; \theta) = \theta/4$ . Uma amostra de  $n = 100$  indivíduos da população apresenta  $n_1 = 65$ ,  $n_2 = 6$ ,  $n_3 = 8$  e  $n_4 = 21$ . Verifique se os dados obtidos suportam o modelo genético acima para a distribuição dos tipos de sangue na população de onde foi selecionada a amostra.

**6.16.** Desenvolva o teste da razão de verossimilhanças generalizada para testar  $H_0 : \beta = \beta_0$  versus  $H_1 : \beta \neq \beta_0$  no modelo de regressão descrito no Exercício 2.12.

**6.17.** O teste  $t$  pareado. Sejam  $(X_1, Y_1), \dots, (X_n, Y_n)$  uma amostra aleatória da variável aleatória bidimensional  $(X, Y)$  com distribuição normal bivariada como dada no Exemplo 2.4.4. Mostre que para testar  $H_0 : \mu_x = \mu_y$  versus  $H_1 : \mu_x \neq \mu_y$ , o teste da razão de verossimilhanças generalizado apresenta região crítica dada por

$$A^* = \left\{ \mathbf{d}; \frac{\sqrt{n}|\bar{d}|}{S_d} > c \right\},$$

onde  $\bar{d} = \sum_{i=1}^n d_i/n$  e  $S_d^2 = \sum_{i=1}^n (d_i - \bar{d})^2/(n-1)$ .



## Referências

1. BICKEL, P.J. e DOKSUM, K.A. (1977). *Mathematical Statistical. Basic Ideas and Selected Topics*. Holden-Day.
2. BUSSAB, W.O. e MORETTIN, P.A. (1987). *Estatística Básica*. São Paulo: Atual.
3. DEGROOT, M.H. (1989). *Probability and Statistics*. New York: Addison-Wesley.
4. FELLER, W. (1976). *Probabilidades*. São Paulo: Edgard Blücher.
5. JAMES, B.R. (1981). *Probabilidade: Um Curso em Nível Intermediário*. Rio de Janeiro: Livro Técnico.
6. LEHMANN, E.L. (1986). *Testing Statistical Hypotheses*. Wiley: New York.
7. SEN, P.K. e SINGER, J.M. (1993). *Large Sample Methods in Statistics. An Introduction with Applications*. Chapman and Hall.

