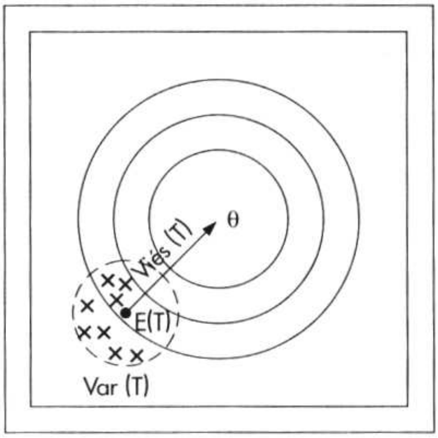
Vemos, portanto, que um estimador preciso tem variância pequena, mas pode ter EQM grande.

Figura 11.2: Representação gráfica para o *EQM*.



Problemas

1. Obtenha a distribuição de $\hat{p}$ quando $p = 0{,}2$ e $n = 5$. Depois calcule $E(\hat{p})$ e $\mathrm{Var}(\hat{p})$.

2. Encontre um limite superior para $\mathrm{Var}(\hat{p})$ quando $n = 10, 25, 100$ e 400. Faça o gráfico em cada caso.

3. Suponha um experimento consistindo de n provas de Bernoulli, com probabilidade de sucesso p. Seja X o número de sucessos, e considere os estimadores

   (a) $\hat{p}_1 = X/n$; (b) $\hat{p}_2 = \begin{cases} 1, & \text{se a primeira prova resultar sucesso,} \\ 0, & \text{caso contrário.} \end{cases}$

   Determine a esperança e a variância de cada estimador. Por que $\hat{p}_2$ não é um "bom" estimador?

4. Verifique se $\hat{p}_1$ e $\hat{p}_2$ do problema 3 são consistentes.

5. Tem-se duas fórmulas distintas para estimar um parâmetro populacional θ. Para ajudar a escolher o melhor, simulou-se uma situação onde $\theta = 100$. Dessa população retiraram-se 1.000 amostras de dez unidades cada uma, e aplicaram-se ambas as fórmulas às dez unidades de cada amostra. Desse modo obtêm-se 1.000 valores para a primeira fórmula t_1 e outros 1.000 valores para a segunda fórmula t_2, cujos estudos descritivos estão resumidos abaixo. Qual das duas fórmulas você acha mais conveniente para estimar θ. Por quê?

Fórmula 1	Fórmula 2
$\bar{t}_1 = 102$ $\mathrm{Var}(t_1) = 5$ Mediana = 100 Moda = 98	$\bar{t}_2 = 100$ $\mathrm{Var}(t_2) = 10$ Mediana = 100 Moda = 100

11.3 Estimadores de Momentos

Neste capítulo e em anteriores, temos usado certos estimadores de parâmetros populacionais, como a média e a variância, simplesmente tentando "imitar" na amostra o que acontece na população. Foi assim que construímos $\overline{X}$, por exemplo.

A média populacional é um caso particular daquilo que chamamos de momento. Na realidade, ela é o *primeiro momento*. Se X for uma v.a. contínua, com densidade $f(x; \theta_1, ..., \theta_r)$, dependendo de r parâmetros, então

$$\mu_1 = E(X) = \int_{-\infty}^{\infty} x f(x; \theta_1, ..., \theta_r)dx. \quad (11.21)$$

Essa média dependerá, genericamente, dos parâmetros desconhecidos $\theta_1, ..., \theta_r$. Por exemplo, suponha que X tenha distribuição normal, com parâmetros μ e σ^2. Aqui, $\theta_1 = \mu$, $\theta_2 = \sigma^2$ e $r = 2$. Temos, nesse caso, que $E(X) = \mu$.

Podemos, em geral, definir o *k-ésimo momento* de X por

$$\mu_k = E(X^k) = \int_{-\infty}^{\infty} x^k f(x; \theta_1, ..., \theta_r)dx, \quad k = 1, 2, ... \quad (11.22)$$

Assim, para $k = 2$, obtemos o segundo momento

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x; \theta_1, ..., \theta_r)dx.$$

No caso acima da normal, temos que $E(X^2) = \text{Var}(X) + [E(X)]^2 = \sigma^2 + \mu^2$. Suponha, agora, que colhemos uma amostra de tamanho n da população $(X_1, ..., X_n)$. Definimos o chamado *k-ésimo momento amostral* por

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \quad k = 1, 2, ... \quad (11.23)$$

Temos, portanto, que $m_1 = \overline{X}$ e $m_2 = \sum_{i=1}^{m} X_i^2/n$.

Definição. Dizemos que $\hat{\theta}_1, ..., \hat{\theta}_r$ são estimadores obtidos pelo método dos momentos se eles forem soluções das equações

$$m_k = \mu_k, \quad k = 1, 2, ..., r. \quad (11.24)$$

O procedimento consiste em substituir os momentos teóricos pelos respectivos momentos amostrais.

Exemplo 11.8. Se $X \sim N(\mu, \sigma^2)$, teremos as seguintes relações válidas para os dois primeiros momentos populacionais:

$$E(X) = \mu, \quad E(X^2) = \sigma^2 + \mu^2,$$

do que obtemos

$$\mu = E(X), \quad \sigma^2 = E(X^2) - E^2(X).$$

Temos, também, os dois primeiros momentos amostrais:

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X},$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Os estimadores obtidos pelo método dos momentos serão

$$\hat{\mu}_M = m_1 = \overline{X},$$

$$\hat{\sigma}_M^2 = m_2 - m_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \hat{\sigma}^2.$$

Ou seja, obtemos os já mencionados estimadores $\overline{X}$ e $\hat{\sigma}^2$.

Na realidade, podemos ter, às vezes, mais de um estimador de momentos. Suponha, por exemplo, que a v.a. Y tenha uma distribuição de Poisson com parâmetro $\lambda > 0$. Vimos que $E(Y) = \text{Var}(Y) = \lambda$, de modo que λ pode ser estimado por $\overline{Y}$ ou por $\sum_{i=1}^{n} (Y_i - \overline{Y})^2/n$, ou seja, $\hat{\lambda}_M = \overline{X}$ ou $\hat{\lambda}_M = \hat{\sigma}^2$, que podem resultar em valores muito diferentes. Veja o Problema 46.

11.4 Estimadores de Mínimos Quadrados

Um dos procedimentos mais usados para obter estimadores é aquele que se baseia no princípio dos mínimos quadrados, introduzido por Gauss em 1794, mas que primeiro apareceu com esse nome no apêndice do tratado de Legendre, *Nouvelles Méthodes pour la Determination des Orbites des Comètes*, publicado em Paris em 1806. Gauss somente viria a publicar seus resultados em 1809, em Hamburgo. Ambos utilizaram o princípio em conexão com problemas de Astronomia e Física.

Vejamos o procedimento por meio de um exemplo simples.

Exemplo 11.9. Um engenheiro está estudando a resistência Y de uma fibra em função de seu diâmetro X e notou que as variáveis são aproximadamente proporcionais, isto é, elas obedecem à relação

$$Y \approx \theta X, \tag{11.25}$$

onde θ é o coeficiente de proporcionalidade. Agora ele deseja estimar o parâmetro θ, baseado numa amostra de cinco unidades, que, submetidas a mensuração e testes, produziram os resultados:

$$X: \quad 1{,}2 \quad 1{,}5 \quad 1{,}7 \quad 2{,}0 \quad 2{,}6, \quad \overline{X} = 1{,}8;$$

$$Y: \quad 3{,}9 \quad 4{,}7 \quad 5{,}6 \quad 5{,}8 \quad 7{,}0, \quad \overline{Y} = 5{,}4.$$

Inspecionando os resultados, conclui-se que $\hat{\theta} = 3$ parece ser um valor razoável. Como verificar a qualidade dessa estimativa? Podemos utilizar o modelo $\hat{Y} = 3X$ e ver como esse prevê os valores de Y, para os dados valores de X, e como são as discrepâncias entre os valores observados e os estimados pelo modelo. Essa análise está resumida na Tabela 11.1.

Os valores da coluna $(Y - 3X)$ medem a inadequação do modelo para cada observação da amostra, enquanto o valor $\sum_{i=1}^{5}(Y_i - 3X_i)^2 = 1{,}06$ é uma tentativa de medir "o erro quadrático total da amostra". Como em situações anteriores, elevou-se ao quadrado para evitar o problema do sinal. Quanto menor for o erro quadrático total, melhor será a estimativa. Isso nos sugere procurar a estimativa que torne mínima essa soma de quadrados. Matematicamente, o problema passa a ser o de encontrar o valor de θ que minimize a função

$$S(\theta) = \sum_{i=1}^{5}(Y_i - \theta X_i)^2. \tag{11.26}$$

Tabela 11.1: Análise do modelo $\hat{Y} = 3X$.

X	Y	$3X$	$Y - 3X$	$(Y - 3X)^2$
1,2	3,9	3,6	0,3	0,09
1,5	4,7	4,5	0,2	0,04
1,7	5,6	5,1	0,5	0,25
2,0	5,8	6,0	−0,2	0,04
2,6	7,0	7,8	0,8	0,64
		Total	0	1,06

O mínimo da função é obtido derivando-a em relação a θ, e igualando o resultado a zero (Ver Morettin et al., 1999), o que resulta

$$\frac{dS(\theta)}{d\theta} = \sum_{i=1}^{5}(Y_i - \hat{\theta}X_i)(-2X_i) = 0.$$

Resolvendo essa equação, obtemos

$$\hat{\theta}_{MQ} = \frac{\sum_{i=1}^{5} X_i Y_i}{\sum_{i=1}^{5} X_i^2}.$$

Usando os dados acima encontramos $\hat{\theta}_{MQ} = 2{,}94$, que conduz a um valor mínimo para $S(\theta)$ de 0,94. Observe que esse valor é realmente menor do que o observado para $\theta = 3$, ou seja, 1,06.

Como foi dito, não esperávamos uma relação perfeita entre as duas variáveis, já que o diâmetro da fibra não é o único responsável pela resistência; outros fatores não controlados afetam o resultado. Desse modo, duas amostras obtidas do mesmo diâmetro X não teriam obrigatoriamente que apresentar o mesmo resultado Y, mas valores em torno de um valor esperado θX.

Em outras palavras, estamos supondo que, para um dado valor da variável explicativa X, os valores da variável resposta Y seguem uma distribuição de probabilidade $f_Y(y)$, centrada em θX. Isso equivale a afirmar que, para cada X, o desvio $\varepsilon = Y - \theta X$ segue uma distribuição centrada no zero. Para melhor entendimento dessa proposição, veja o Capítulo 16. Podemos, então, escrever

$$E(Y|x) = \theta x, \quad \text{para todo valor } x.$$

É comum supor que ε tem a mesma distribuição, para todo valor x da variável explicativa X. Desse modo, é comum escrever

$$Y = \theta x + \varepsilon,$$

com ε seguindo a distribuição $f_\varepsilon(.)$, com média zero. Como ilustração, poderíamos supor que $\varepsilon \sim N(0, \sigma^2)$, para todo x. Quanto menor for a variância σ^2, melhor será a "previsão" de Y como função de x. Assim, parece razoável escolher θ que torna mínima a soma dos quadrados do erros:

$$\sum_{i=1}^{5} \varepsilon_i^2 = \sum_{i=1}^{5} (Y_i - \theta X_i)^2.$$

O modelo acima pode ser generalizado, de modo a envolver outras funções do parâmetro θ, resultando no modelo

$$Y = g(X; \theta) + \varepsilon, \tag{11.27}$$

e devemos procurar o valor de θ que minimize a função

$$S(\theta) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - g(X_i; \theta))^2, \tag{11.28}$$

para uma amostra $(X_1, Y_1), ..., (X_n, Y_n)$ das variáveis X e Y. A solução $\hat{\theta}_{MQ}$ é chamada de estimador de mínimos quadrados (EMQ) de θ.

No Capítulo 16 voltaremos a esse tópico e trataremos com mais detalhes o chamado modelo de regressão linear simples.

Problemas

6. Estamos estudando o modelo $y_t = \mu + a_t$, para o qual uma amostra de cinco elementos produziu os seguintes valores para y_t: 3, 5, 6, 8, 16.
   (a) Calcule os valores de $S(\mu) = \sum_t (y_t - \mu)^2$, para $\mu = 6, 7, 8, 9, 10$, e faça o gráfico de $S(\mu)$ em relação a μ. Qual o valor de μ que parece tornar mínimo $S(\mu)$?
   (b) Derivando $S(\mu)$ em relação a μ, e igualando o resultado a zero, você encontrará o EMQ de μ. Usando os dados acima, encontre a estimativa para μ e compare com o resultado do item anterior.

7. Os dados abaixo referem-se ao índice de inflação (y_t) de 1967 a 1979.

Ano (t)	1967	1969	1971	1973	1975	1977	1979
Inflação (y_t)	128	192	277	373	613	1.236	2.639

   (a) Faça o gráfico de y_t contra t.
   (b) Considere ajustar o modelo $y_t = \alpha + \beta t + \varepsilon_t$ aos dados. Encontre as estimativas de mínimos quadrados de α e β.
   (c) Qual seria a inflação em 1981?
   (d) Você teria alguma restrição em adotar o modelo linear nesse caso?

8. No problema 7, determinamos os estimadores de mínimos quadrados para o modelo $y_t = f(t) + \epsilon_t$, no qual $f(t) = \alpha + \beta t$. Suponha agora que

$$f(t) = \alpha + \beta x_t, \quad t = 1, ..., n,$$

ou seja, temos n valores fixos $x_1, ..., x_n$ de uma variável fixa (não-aleatória) x. Obtenha os EMQ de α e β para esse modelo.

9. Aplique os resultados do problema 8 para os dados a seguir:

t	1	2	3	4	5	6	7	8	9	10
x_t	1,5	1,8	1,6	2,5	4,0	3,8	4,5	5,1	6,5	6,0
y_t	66,8	67,0	66,9	67,6	68,9	68,7	69,3	69,8	71,0	70,6

11.5 Estimadores de Máxima Verossimilhança

O Novo Dicionário Aurélio da Língua Portuguesa (2ª edição, 1986) define *verossímil* (ou verossimilhante) aquilo que é semelhante à verdade, provável, e *verossimilhança* (ou verossimilidade, ou ainda verossimilitude), à qualidade ou caráter de verossímil. O que seria uma amostra verossímil? Seria uma amostra que fornecesse a melhor informação possível sobre um parâmetro de interesse da população, desconhecido, e que desejamos estimar.

O princípio da verossimilhança afirma que devemos escolher aquele valor do parâmetro desconhecido que maximiza a probabilidade de obter a amostra particular observada, ou seja, o valor que torna aquela amostra a "mais provável". O uso desse princípio conduz a um método de estimação pelo qual se obtêm os chamados estimadores de máxima verossimilhança que, em geral, têm propriedades muito boas. Esse princípio foi enunciado por Fisher pela primeira vez em 1912 e, em 1922, deu-lhe forma mais completa, introduzindo a expressão "likelihood" (verossimilhança). Veja Fisher (1935) para mais detalhes. Vamos começar com um exemplo.

Exemplo 11.10. Suponha que temos n provas de Bernoulli com P (sucesso) $= p$, $0 < p < 1$ e X = número de sucessos. Devemos tomar como estimador aquele valor de p que torna a amostra observada a mais provável de ocorrer.

Suponha, por exemplo, que $n = 3$ e obtemos dois sucessos e um fracasso. A função de verossimilhança é

$$L(p) = P(2 \text{ sucessos e } 1 \text{ fracasso}) = p^2(1 - p).$$

Maximizando essa função em relação a p, obtemos

$$L'(p) = 2p(1 - p) - p^2 = 0 \Rightarrow p(2 - 3p) = 0,$$

do que seguem $p = 0$ ou $p = 2/3$. É fácil ver que o ponto máximo é $\hat{p} = 2/3$, que é o estimador de máxima verossimilhança (EMV) de p.

De modo geral, o EMV do parâmetro p de uma distribuição binomial é

$$\hat{p}_{MV} = \frac{X}{n}, \tag{11.29}$$

que é o estimador usado anteriormente no Exemplo 11.1.

Para chegar a (11.29), observe que a função de verossimilhança nesse caso é

$$L(p) = p^x (1-p)^{n-x},$$

que é a probabilidade de se obter x sucessos e $n - x$ fracassos. O máximo dessa função ocorre no mesmo ponto que $\ell(p) = \log_e L(p)$. Denotando o logaritmo natural simplesmente por log, temos

$$\ell(p) = x \log p + (n-x)\log(1-p).$$

Derivando e igualando a zero obtemos $\hat{p}_{MV} = x/n$.

O procedimento, pois, é obter a função de verossimilhança, que depende dos parâmetros desconhecidos e dos valores amostrais, e depois maximizar essa função ou o logaritmo dela, o que pode ser mais conveniente em determinadas situações. Chamando de $L(\theta; X_1, ..., X_n)$ a função de verossimilhança, a log-verossimilhança será $\ell(\theta; X_1, ..., X_n) = \log_e L(\theta; X_1, ..., X_n)$.

No caso de variáveis contínuas, a função de verossimilhança é definida da seguinte maneira. Suponha que a v.a. X tenha densidade $f(x; \theta)$, onde destacamos a dependência do parâmetro θ desconhecido. Retiramos uma amostra de X, de tamanho n, $(X_1, ..., X_n)$, e sejam $(x_1, ..., x_n)$ os valores efetivamente observados.

Definição. A função de verossimilhança é definida por

$$L(\theta; x_1, ..., x_n) = f(x_1; \theta) \dots f(x_n; \theta), \tag{11.30}$$

que deve ser encarada como uma função de θ. O estimador de máxima verossimilhança de θ é o valor $\hat{\theta}_{MV}$ que maximiza $L(\theta; x_1, ..., x_n)$.

Se indicarmos por $\boldsymbol{x} = (x_1, ..., x_n)'$ o vetor contendo a amostra, é costume denotar a verossimilhança por $L(\theta|\boldsymbol{x})$ e a log-verossimilhança por $\ell(\theta|\boldsymbol{x})$. O parâmetro θ pode ser um vetor, como no caso de querermos estimar a média μ e a variância σ^2 de uma normal. Nesse caso, $\theta = (\mu, \sigma^2)'$.

Exemplo 11.11. Suponha que a v.a. X tenha distribuição exponencial, com parâmetro $\alpha > 0$, desconhecido, e queremos obter o EMV desse parâmetro. A densidade de X é dada por (7.26):

$$f(x; \alpha) = \begin{cases} \frac{1}{\alpha} e^{-x/\alpha}, & \text{se } x \geqslant 0 \\ 0, & \text{se } x < 0. \end{cases}$$

Então, a verossimilhança é dada por

$$L(\alpha|x) = (1/\alpha)^n e^{-\sum_i x_i/\alpha}$$

e a log-verossimilhança fica

$$\ell(\alpha|x) = -n \log \alpha - \sum_{i=1}^{n} x_i/\alpha.$$

Derivando e igualando a zero obtemos que o EMV de α é

$$\hat{\alpha}_{MV} = \frac{\sum_{i=1}^{n} x_i}{n}, \tag{11.31}$$

que nada mais é do que a média amostral. Lembremos que na distribuição exponencial E(T) = α, e portanto o estimador obtido é o esperado pelo senso comum.

No caso discreto, a função de verossimilhança pode ser escrita na forma

$$L(\theta; x_1, ..., x_n) = P(X_1 = x_1|\theta) \ldots P(X_n = x_n|\theta).$$

Veja o problema 37 para o caso de termos mais de um parâmetro.

Problemas

10. Na função de verossimilhança $L(p)$ da binomial, suponha que $n = 5$ e $x = 3$. Construa o gráfico da função para os possíveis valores de p = 1/5, 2/5, 3/5, 4/5, e verifique que o máximo ocorre realmente para p = 3/5.

11. Observa-se uma seqüência de ensaios de Bernoulli, independentes, com parâmetro p, até a ocorrência do primeiro sucesso. Se X indicar o número de ensaios necessários:
    (a) Mostre que $P(X = x) = (1 - p)^{x-1}p$ (distribuição geométrica).
    (b) Repetiu-se esse experimento n vezes, e em cada um deles o número de ensaios necessários foram $x_1, x_2, ..., x_n$. Encontre o EMV para p.
    (c) Usando uma moeda, repetiu-se esse experimento 5 vezes, e o número de ensaios necessários até a ocorrência da primeira coroa foi 2, 3, 1, 4, 1, respectivamente. Qual a estimativa de MV para p = probabilidade de ocorrência de coroa nessa moeda? Existiria outra maneira de estimar p?

12. Suponha que X seja uma v.a. com distribuição normal, com média μ e variância 1. Obtenha o EMV de μ, para uma amostra de tamanho n, $(x_1, ..., x_n)$.

13. Considere Y uma v.a. com distribuição de Poisson, com parâmetro $\lambda > 0$. Obtenha a EMV de λ, baseado numa amostra de tamanho n.

11.6 Intervalos de Confiança

Até agora, todos os estimadores apresentados foram pontuais, isto é, especificam um único valor para o estimador. Esse procedimento não permite julgar qual a possível magnitude do erro que estamos cometendo. Daí, surge a idéia de construir os intervalos de confiança, que são baseados na distribuição amostral do estimador pontual.

Exemplo 11.12. Suponha que queiramos estimar a média μ de uma população qualquer, e para tanto usamos a média $\overline{X}$ de uma amostra de tamanho n. Do TLC,

$$e = (\overline{X} - \mu) \sim N(0, \sigma^2_{\bar{x}}), \qquad (11.32)$$

com $\mathrm{Var}(\overline{X}) = \sigma^2_{\bar{x}} = \sigma^2/n$. Daqui podemos determinar qual a probabilidade de cometermos erros de determinadas magnitudes. Por exemplo,

$$P(|e| < 1{,}96\sigma_{\bar{x}}) = 0{,}95$$

ou

$$P(|\overline{X} - \mu| < 1{,}96\sigma_{\bar{x}}) = 0{,}95,$$

que é equivalente a

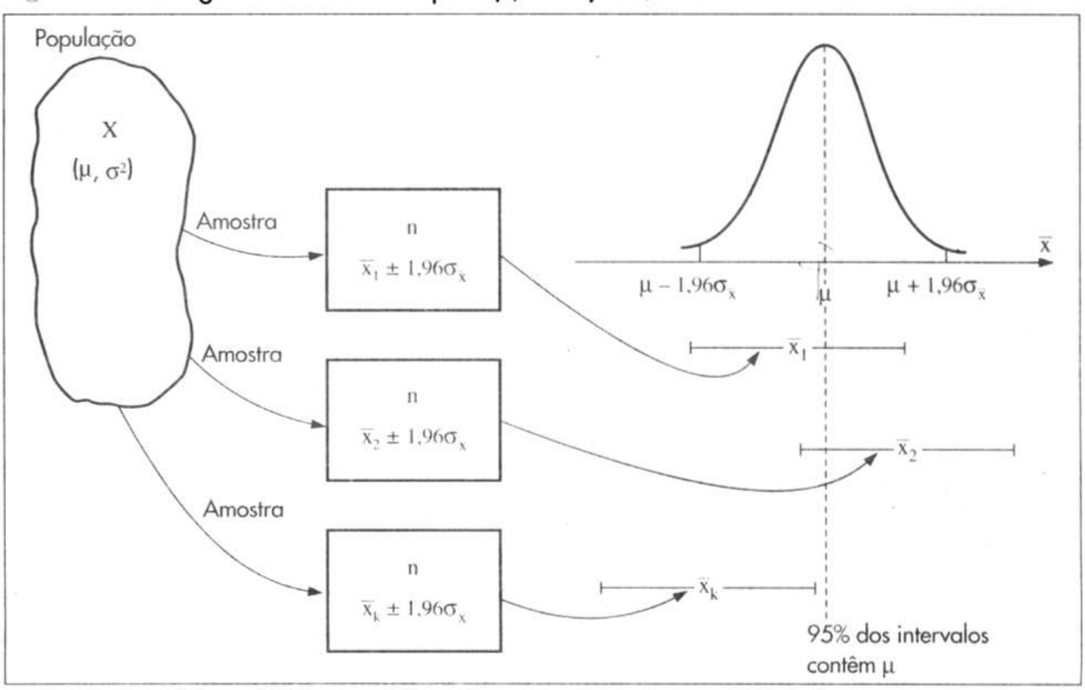
$$P(-1{,}96\sigma_{\bar{x}} < \overline{X} - \mu < 1{,}96\sigma_{\bar{x}}) = 0{,}95,$$

e, finalmente,

$$P(\overline{X} - 1{,}96\sigma_{\bar{x}} < \mu < \overline{X} + 1{,}96\sigma_{\bar{x}}) = 0{,}95. \qquad (11.33)$$

Convém lembrar que μ não é uma variável aleatória e sim um parâmetro, e a expressão (11.33) deve ser interpretada da seguinte maneira: se pudéssemos construir uma quantidade grande de intervalos (aleatórios!) da forma $]\overline{X} - 1{,}96\sigma_{\bar{x}},\ \overline{X} + 1{,}96\sigma_{\bar{x}}[$, todos baseados em amostras de tamanho n, 95% deles conteriam o parâmetro μ. Veja a Figura 11.3. Dizemos que $\gamma = 0{,}95$ é o *coeficiente de confiança*. Nessa figura estão esquematizados o funcionamento e o significado de um intervalo de confiança (IC) para μ, com $\gamma = 0{,}95$ e σ^2 conhecido.

Figura 11.3: Significado de um IC para μ, com $\gamma = 0{,}95$ e σ^2 conhecido.



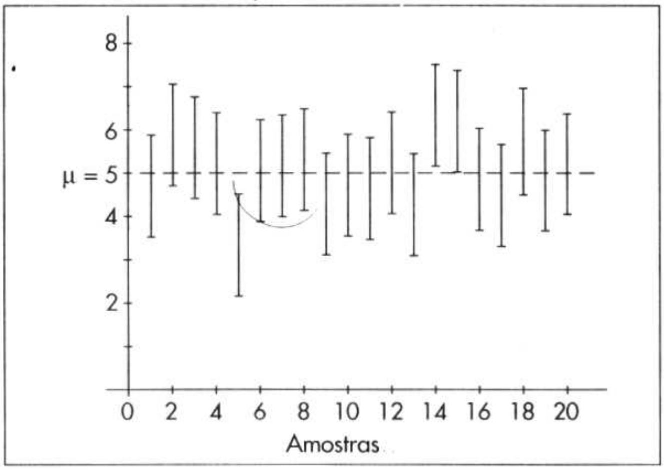
Escolhida uma amostra e encontrada sua média $\bar{x}_0$, e admitindo-se $\sigma_{\bar{x}}$ conhecido, podemos construir o intervalo

$$]\bar{x}_0 - 1{,}96\sigma_{\bar{x}}, \bar{x}_0 + 1{,}96\sigma_{\bar{x}}[. \qquad (11.34)$$

Esse intervalo pode ou não conter o parâmetro μ, mas pelo exposto acima temos 95% de confiança de que contenha.

Para ilustrar o que foi dito acima, consideremos o seguinte experimento de simulação. Geramos 20 amostras de tamanho $n = 25$ de uma distribuição normal de média $\mu = 5$ e desvio padrão $\sigma = 3$. Para cada amostra construímos o intervalo de confiança para μ, com coeficiente de confiança $\gamma = 0{,}95$, que é da forma $\bar{X} \pm 1{,}176$, usando (11.34). Na Figura 11.4, temos esses intervalos representados e notamos que três deles (amostras de números 5, 14 e 15) não contêm a média $\mu = 5$.

Figura 11.4: Intervalos de confiança para a média de uma N(5, 9), para 20 amostras de tamanho n = 25.



Exemplo 11.13. Uma máquina enche pacotes de café com uma variância igual a 100 g². Ela estava regulada para encher os pacotes com 500 g, em média. Agora, ela se desregulou, e queremos saber qual a nova média μ. Uma amostra de 25 pacotes apresentou uma média igual a 485 g. Vamos construir um intervalo de confiança com 95% de confiança para μ. De (11.34), teremos

$$IC(\mu; 0{,}95) = 485 \pm 1{,}96 \times 2,$$

ou seja,

$$IC(\mu; 0{,}95) =]481, 489[,$$

pois $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 10/5 = 2$g.

Se T for um estimador do parâmetro θ, e conhecida a distribuição amostral de T, sempre será possível achar dois valores t_1 e t_2, tais que

$$P(t_1 < \theta < t_2) = \gamma, \qquad (11.35)$$

a probabilidade interpretada como em (11.33), e γ um valor fixo, $0 < \gamma < 1$. Para uma dada amostra, teremos dois valores fixos para t_1 e t_2, e o intervalo de confiança para θ, com coeficiente de confiança γ, será indicado do seguinte modo:

$$\mathrm{IC}(\theta;\ \gamma) =]t_1,\ t_2[. \tag{11.36}$$

Se a variância populacional σ^2 não for conhecida, podemos substituir em (11.34) $\sigma_{\bar{x}}$ por $S/\sqrt{n}$, onde S^2 é a variância amostral dada em (11.9). Para n grande, da ordem de 100, o intervalo (11.34), com essa modificação, pode ainda ser usado. Para n não muito grande, a distribuição normal não pode mais ser usada e terá de ser substituída pela distribuição t de Student, que estudamos no Capítulo 7. Esse assunto voltará a ser abordado no Capítulo 12.

Para um coeficiente de confiança qualquer γ, teremos de usar o valor $z(\gamma)$ tal que $P(-z(\gamma) < Z < z(\gamma)) = \gamma$, com $Z \sim N(0,\ 1)$. O intervalo fica

$$\mathrm{IC}(\mu;\ \gamma) =]\bar{X} - z(\gamma)\sigma_{\bar{x}};\ \bar{X} + z(\gamma)\sigma_{\bar{x}}[. \tag{11.37}$$

Observe, também, que a amplitude do intervalo (11.37) é $L = 2z(\gamma)\sigma/\sqrt{n}$, que é uma constante, independente de $\bar{X}$. Se construirmos vários intervalos de confiança para o mesmo valor de n, σ e γ, estes terão extremos aleatórios, mas todos terão a mesma amplitude L.

Exemplo 11.14. Vamos obter um intervalo de confiança para o parâmetro p de uma distribuição $b(n, p)$. Sabemos que se X = número de sucessos nas n provas, então X tem distribuição aproximadamente normal, com média $\mu = np$ e variância $\sigma^2 = npq$, com $q = 1 - p$. Logo,

$$Z = \frac{X - np}{\sqrt{npq}} \sim N(0,\ 1),$$

ou ainda,

$$Z = \frac{X/n - p}{\sqrt{pq/n}} = \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{pq}} \sim N(0,\ 1). \tag{11.38}$$

Assim, se $\gamma = 0{,}95$, temos, consultando a Tabela III, que

$$P(-1{,}96 \leq Z \leq 1{,}96) = 0{,}95,$$

ou seja,

$$P\left\{-1{,}96 \leq \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{pq}} \leq 1{,}96\right\} = 0{,}95.$$

Portanto, com probabilidade 0,95, temos que

$$-1{,}96\sqrt{pq/n} \leq \hat{p} - p \leq 1{,}96\sqrt{pq/n},$$

do que segue

$$\hat{p} - 1{,}96\sqrt{pq/n} \leq p \leq \hat{p} + 1{,}96\sqrt{pq/n}.$$

Como não conhecemos p, podemos proceder de duas maneiras. Uma é usar o fato que $pq \leqslant 1/4$, de modo que $\sqrt{pq/n} \leqslant 1/\sqrt{4n}$, obtendo

$$\hat{p} - \frac{1{,}96}{\sqrt{4n}} \leqslant p \leqslant \hat{p} + \frac{1{,}96}{\sqrt{4n}}. \qquad (11.39)$$

Temos, então, que $]\hat{p} - 1{,}96/\sqrt{4n}; \hat{p} + 1{,}96/\sqrt{4n}[$ é um intervalo de confiança para p, com coeficiente de confiança de 95%.

Para um γ qualquer, $0 < \gamma < 1$, (11.39) fica

$$\hat{p} - \frac{z(\gamma)}{\sqrt{4n}} \leqslant p \leqslant \hat{p} + \frac{z(\gamma)}{\sqrt{4n}}, \qquad (11.40)$$

onde $z(\gamma)$ é definido como em (11.37).

Exemplo 11.15. Numa pesquisa de mercado, $n = 400$ pessoas foram entrevistadas sobre determinado produto, e 60% delas preferiram a marca A. Aqui, $\hat{p} = 0{,}6$ e um intervalo de confiança para p com coeficiente de confiança $\gamma = 0{,}95$ será

$$0{,}6 \pm (1{,}96)\ 1/\sqrt{1\,600} = 0{,}6 \pm 0{,}049,$$

ou seja

$$IC(p; 0{,}95) =]0{,}551; 0{,}649[.$$

O intervalo (11.40) é chamado *conservador*, pois se p não for igual a 1/2 e estiver próximo de zero ou de um, então ele fornece um intervalo desnecessariamente maior, porque substituímos pq pelo seu valor máximo, 1/4. Uma outra maneira de proceder é substituir pq por $\hat{p}\hat{q}$, com $\hat{q} = 1 - \hat{p}$. O intervalo obtido fica

$$\hat{p} - z(\gamma)\sqrt{\hat{p}\hat{q}/n} \leqslant p \leqslant \hat{p} + z(\gamma)\sqrt{\hat{p}\hat{q}/n}, \qquad (11.41)$$

com $z(\gamma)$ definido como em (11.40).

Exemplo 11.16. Suponha que em $n = 400$ provas obtemos $k = 80$ sucessos. Vamos obter um intervalo de confiança para p com $\gamma = 0{,}90$. Como $\hat{p} = 80/400 = 0{,}2$ e $\hat{q} = 1 - \hat{p} = 0{,}8$, então (11.41) fica

$$0{,}2 \pm (1{,}645)\sqrt{(0{,}2)(0{,}8)/400} = 0{,}2 \pm 0{,}033,$$

ou seja,

$$IC(p; 0{,}90) =]0{,}167; 0{,}233[.$$

Usando (11.40) o intervalo conservador é

$$IC(p; 0{,}90) =]0{,}159; 0{,}241[.$$

Observe que o primeiro intervalo tem amplitude menor que o segundo. Outra observação importante é que por (11.40) e um γ fixo, os intervalos que podemos obter para amostras diferentes (mas de mesmo tamanho n) terão a mesma amplitude, dada por $2z(\gamma)/\sqrt{4n}$. Por outro lado, usando (11.41), a amplitude do intervalo será $2z(\gamma)\frac{\sqrt{\hat{p}\hat{q}}}{n}$, que é variável de amostra para amostra, pois $\hat{p}$ (e, conseqüentemente, $\hat{q}$) variará de amostra para amostra.

Problemas

14. Calcule o intervalo de confiança para a média de uma $N(\mu, \sigma^2)$ em cada um dos casos abaixo.

Média Amostral	Tamanho da Amostra	Desvio Padrão da População	Coeficiente de Confiança
170 cm	100	15 cm	95%
165 cm	184	30 cm	85%
180 cm	225	30 cm	70%

15. De 50.000 válvulas fabricadas por uma companhia retira-se uma amostra de 400 válvulas, e obtém-se a vida média de 800 horas e o desvio padrão de 100 horas.
    (a) Qual o intervalo de confiança de 99% para a vida média da população?
    (b) Com que confiança dir-se-ia que a vida média é $800 \pm 0{,}98$?
    (c) Que tamanho deve ter a amostra para que seja de 95% a confiança na estimativa $800 \pm 7{,}84$?
    (Que suposições você fez para responder às questões acima?)

16. Qual deve ser o tamanho de uma amostra cujo desvio padrão é 10 para que a diferença da média amostral para a média da população, em valor absoluto, seja menor que 1, com coeficiente de confiança igual a:
    (a) 95% (b) 99%

17. Uma população tem desvio padrão igual a 10.
    (a) Que tamanho deve ter uma amostra para que, com probabilidade 8%, o erro em estimar a média seja superior a uma unidade?
    (b) Supondo-se colhida a amostra no caso anterior, qual o intervalo de confiança, se $\bar{x} = 50$?

18. Uma amostra aleatória de 625 donas de casa revela que 70% delas preferem a marca A de detergente. Construir um intervalo de confiança para p = proporção das donas de casa que preferem A com c.c. $\gamma = 90\%$.

19. Encontre os intervalos de confiança para p se $k/n = 0{,}3$, com c.c. $\gamma = 0{,}95$. Utilize os dois enfoques apontados na seção 11.6, com $n = 400$.

20. Antes de uma eleição, um determinado partido está interessado em estimar a proporção p de eleitores favoráveis ao seu candidato. Uma amostra piloto de tamanho 100 revelou que 60% dos eleitores eram favoráveis ao candidato em questão.
    (a) Determine o tamanho da amostra necessário para que o erro cometido na estimação seja de, no máximo, 0,01 com probabilidade de 80%.
    (b) Se na amostra final, com tamanho igual ao obtido em (a), observou-se que 55% dos eleitores eram favoráveis ao candidato em questão, construa um intervalo de confiança para a proporção p. Utilize $\gamma = 0{,}95$.

21. Suponha que estejamos interessados em estimar a porcentagem de consumidores de um certo produto. Se a amostra de tamanho 300 forneceu 100 indivíduos que consomem o dado produto, determine:
    (a) o intervalo de confiança de p, com coeficiente de confiança de 95% (interprete o resultado);
    (b) o tamanho da amostra para que o erro da estimativa não exceda a 0,02 unidades com probabilidade de 95% (interprete o resultado).

11.7 Erro Padrão de um Estimador

Vimos que, obtida a distribuição amostral de um estimador, podíamos calcular a sua variância. Se não pudermos obter a distribuição exata, usamos uma aproximação, se essa estiver disponível, como no caso de $\bar{X}$, e a variância do estimador será a variância dessa aproximação. Por exemplo, para a média amostral $\bar{X}$, obtida de uma amostra de tamanho n, temos que

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n},$$

na qual σ^2 é a variância da v.a. X definida sobre a população.

À raiz quadrada dessa variância chamaremos de erro padrão de $\bar{X}$ e o denotaremos por

$$\text{EP}(\bar{X}) = \frac{\sigma}{\sqrt{n}}. \tag{11.42}$$

Definição. Se T for um estimador do parâmetro θ, chamaremos de *erro padrão de T* a quantidade

$$\text{EP}(T) = \sqrt{\text{Var}(T)}. \tag{11.43}$$

A variância de T dependerá dos parâmetros da distribuição de X, o mesmo acontecendo com o erro padrão. Por exemplo, em (11.42), $\text{EP}(\bar{X})$ depende de σ, que em geral é desconhecida. Podemos, então, obter o erro padrão estimado de $\bar{X}$, dado por

$$\text{ep}(\bar{X}) = \widehat{\text{EP}}(\bar{X}) = S/\sqrt{n}, \tag{11.44}$$

na qual S^2 é a variância amostral. Genericamente, o *erro padrão estimado* de T é dado por

$$\widehat{\text{EP}}(T) = \sqrt{\widehat{\text{Var}}(T)}. \tag{11.45}$$

Muitas vezes a quantidade (11.45) é chamada de erro amostral. Mas preferimos chamar de erro amostral à diferença $e = T - \theta$.

Exemplo 11.17. Para o Exemplo 11.15, $\hat{p} = 0{,}6$, e o erro padrão de $\hat{p}$ será dado por

$$\text{EP}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}. \qquad (11.46)$$

Como não conhecemos p usamos no seu lugar o estimador $\hat{p}$, obtendo-se

$$\widehat{\text{EP}}(\hat{p}) = \sqrt{(0{,}6)(0{,}4)/400} = 0{,}025.$$

Observe que o intervalo de confiança (11.41) pode ser escrito

$$\hat{p} \pm z(\gamma)(\widehat{\text{EP}}(\hat{p})),$$

ao passo que o intervalo para μ dado por (11.37) pode ser escrito

$$\overline{X} \pm (1{,}96)(\text{EP}(\overline{X})).$$

11.8 Inferência Bayesiana

O estabelecimento de uma ponte entre os valores observados na amostra e os modelos postulados para a população, objeto da inferência estatística, exige a adoção de princípios teóricos muito bem especificados. Neste livro usaremos a chamada *teoria freqüentista* (às vezes também chamada de clássica). Seus fundamentos jazem em trabalhos de J. Neyman, E. Pearson, R. Fisher e outros.

Consideremos um exemplo para ilustrar esse enfoque. Suponha que tenhamos uma amostra observada $(x_1, ..., x_n)$ de uma população normal, $N(\mu, \sigma^2)$, e queremos fazer inferências sobre os valores de μ e σ^2, baseados nas n observações.

Por meio de algum procedimento estudado neste capítulo, selecionamos estimadores $\hat{\mu}(x)$ e $\hat{\sigma}^2(x)$ que sejam funções do vetor de observações $\boldsymbol{x} = (x_1, ..., x_n)'$. Considere dados hipotéticos $\boldsymbol{x}_1, \boldsymbol{x}_2, ...$, todos amostras de tamanho n, que poderiam ter sido gerados da população em questão. Obtemos, então, as distribuições amostrais de $\hat{\mu}(\boldsymbol{x})$ e $\hat{\sigma}^2(\boldsymbol{x})$, como na seção 10.7. Podemos também obter intervalos de confiança para os parâmetros desconhecidos μ e σ^2, bem como testar hipóteses sobre esses parâmetros, assunto a ser discutido no Capítulo 12.

Para construir intervalos de confiança e testar hipóteses será necessário conhecer a distribuição amostral dos estimadores. Como só temos um conjunto de dados e não dados hipotéticos, estas distribuições amostrais terão de ser obtidas de outra maneira, e não como no Exemplo 10.7. Usualmente isso é feito usando teoremas como o Teorema Limite Central, discutido na seção 10.8, obtendo-se uma distribuição aproximada para os estimadores, que vale para tamanhos de amostras grandes.

A crítica que se faz à teoria freqüentista é a possibilidade de "replicar dados", bem como o recurso à teoria assintótica. Uma teoria que não faz uso de tais argumentos é a inferência bayesiana, cujos fundamentos foram estabelecidos por T. Bayes em 1763. Outros expoentes dessa corrente foram Bernoulli (1713), Laplace (1812) e Jeffreys (1939). Aqui, o Teorema de Bayes, estudado no Capítulo 5, tem papel fundamental. A noção de probabilidade prevalente aqui é a subjetiva, discutida brevemente no mesmo capítulo.

Com relação ao nosso exemplo, a Inferência Bayesiana admite que os parâmetros μ e σ^2, que são quantidades desconhecidas da distribuição de X, podem ser descritos por uma distribuição de probabilidades, $p(\mu, \sigma^2)$, chamada a *distribuição a priori* desses parâmetros. Nessa distribuição são incorporadas todas as informações que temos sobre $\theta = (\mu, \sigma^2)'$, inclusive de natureza subjetiva. Essa distribuição é hipotetizada antes de se colherem os dados.

O que é importante observar é que, tanto na teoria freqüentista como na bayesiana, um parâmetro qualquer, como μ, no exemplo acima, é considerado fixo. O que se faz no enfoque bayesiano é caracterizar a incerteza sobre esse parâmetro por meio de uma distribuição de probabilidades.

Após obtidos os dados, obtemos a função de verossimilhança, que incorpora a informação sobre θ fornecida pelos dados. Finalmente, obtemos a distribuição *a posteriori* de θ, dada a amostra observada. Um estimador de θ pode ser tomado, por exemplo, como a média ou a moda dessa distribuição *a posteriori*.

Vimos no Capítulo 5 que o teorema de Bayes pode ser usado para atualizar probabilidades de um evento. Mas o teorema também pode ser utilizado para obter informação sobre um parâmetro desconhecido de um modelo probabilístico, como o binomial ou normal, por exemplo. Chamemos de θ um tal parâmetro, suposto desconhecido, e para o qual tenhamos alguma informação anterior, consubstanciada numa distribuição de probabilidades $p(\theta)$, chamada distribuição *a priori* de θ. Vamos supor, por ser mais simples, que θ tenha os valores $\theta_1, \theta_2, ..., \theta_r$, com probabilidades *a priori* $P(\theta = \theta_i) = p(\theta_i)$, $i = 1, 2, ..., r$. Chamemos de y a nova informação sobre θ, que também é obtida de um modelo discreto. Então o teorema de Bayes pode ser escrito

$$P(\theta_i|y) = \frac{p(\theta_i)P(y\,|\,\theta_i)}{\sum_{j=1}^{r} p(\theta_j)P(y\,|\,\theta_j)}, \quad i = 1, 2, ..., r. \tag{11.47}$$

Aqui, as verossimilhanças são $P(y\,|\,\theta_1), ..., P(y\,|\,\theta_r)$, e as probabilidades *a posteriori* determinadas pelo teorema de Bayes são $P(\theta_1|y), ..., P(\theta_r|y)$. Obtida essa distribuição *a posteriori* de θ, dada a nova informação y, podemos por exemplo estimar θ como sendo a média dessa distribuição ou a moda (o valor que maximiza $P(\theta|y)$).

Exemplo 11.18. Vamos considerar uma aplicação do Teorema de Bayes a um exemplo simples de mercado de ações. Chamemos de y o rendimento do IBOVESPA (Índice da Bolsa de Valores de São Paulo), em porcentagem, por período (mês, por exemplo). Suponha que estejamos interessados somente se o rendimento for positivo ($y > 0$) ou negativo ($y < 0$). Designando por θ o "estado do mercado", vamos considerar apenas dois estados, mercado em alta (θ_1) ou mercado em baixa (θ_2). Suponha que se tenha a seguinte informação prévia (ou *a priori*) sobre as probabilidades de θ_1 e θ_2:

priori	θ_1	θ_2
$p(\theta)$	3/5	2/5

Então, as probabilidades *a priori* dos estados são $p(\theta_1) = P(\theta = \theta_1) = 3/5$ e $p(\theta_2) =$ $= P(\theta = \theta_2) = 2/5$. As verossimilhanças são dadas aqui por

$$P(y > 0|\theta) \quad \text{e} \quad P(y < 0|\theta),$$

para $\theta = \theta_1, \theta_2$, que denotaremos genericamente por $p(y|\theta)$. Essas verossimilhanças são supostas conhecidas no Teorema de Bayes e vamos supor que em nosso caso são dadas na tabela abaixo.

y \ θ	$p(y\|\theta)$ θ_1	θ_2
$y > 0$	2/3	1/3
$y < 0$	1/3	2/3

Ou seja, temos que

$$P(y > 0|\theta_1) = 2/3, \quad P(y > 0|\theta_2) = 1/3,$$
$$P(y < 0|\theta_1) = 1/3, \quad P(y < 0|\theta_2) = 2/3.$$

Podemos calcular as probabilidades conjuntas $p(y, \theta)$, ou seja,

$$p(\mathrm{y}, \theta) = p(\theta)p(y|\theta),$$

obtendo-se a tabela abaixo.

y \ θ	$p(y, \theta)$ θ_1	θ_2	$p(y)$
$y > 0$	6/15	2/15	8/15
$y < 0$	3/15	4/15	7/15
$p(\theta)$	9/15	6/15	1

Por exemplo,

$$P(y > 0, \theta = \theta_1) = P(\theta = \theta_1) \cdot P(y > 0|\theta = \theta_1) = 3/5 \times 2/3 = 6/15.$$

O Teorema de Bayes, dado pela fórmula (11.47), fornece as probabilidades *a posteriori* de θ_1 e θ_2, dado o valor observado de y:

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}. \tag{11.48}$$

Para calcular (11.48) precisamos calcular $p(y)$, que são chamadas probabilidades marginais preditoras ou simplesmente *previsões*. Usando o mesmo argumento que deu origem a (5.14), podemos escrever

$$p(y) = \sum_\theta p(y, \theta) = \sum_\theta p(\theta)p(y|\theta).$$

Em nosso caso,

$$\begin{aligned} P(y > 0) &= P(\theta_1)P(y > 0|\theta_1) + P(\theta_2)P(y > 0|\theta_2) \\ &= 3/5 \times 2/3 + 2/5 \times 1/3 = 8/15. \end{aligned}$$

Do mesmo modo,

$$P(y < 0) = P(\theta_1)P(y < 0|\theta_1) + P(\theta_2)P(y < 0|\theta_2) = 7/15,$$

e teremos a tabela a seguir:

y	$p(y)$
$y > 0$	8/15
$y < 0$	7/15

Vemos que essa é a mesma distribuição marginal de y, dada na tabela que mostra a distribuição conjunta de y e θ.

Então, por (11.48),

$$P(\theta = \theta_1|y > 0) = \frac{P(\theta_1)P(y > 0|\theta_1)}{P(y > 0)} = \frac{3/5 \times 2/3}{8/15} = 3/4,$$

$$P(\theta = \theta_2|y > 0) = \frac{P(\theta_2)P(y > 0|\theta_2)}{P(y > 0)} = 1/4.$$

De modo análogo, obtemos

$$P(\theta = \theta_1|y < 0) = 3/7, \quad P(\theta = \theta_2|y < 0) = 4/7.$$

Temos, então, as probabilidades condicionais de alta e baixa, dada a informação de que o retorno é positivo ou negativo:

y \ θ	$p(\theta\|y)$	
	θ_1	θ_2
$y > 0$	3/4	1/4
$y < 0$	3/7	4/7

Podemos, por exemplo, "estimar" θ (alta ou baixa) por θ_1 (mercado em alta) se $y > 0$, já que $P(\theta = \theta_1|y > 0) = 3/4$ e "estimar" θ por θ_2 (mercado em baixa) se $y < 0$, pois $P(\theta = \theta_2|y < 0) = 4/7$. Ou seja, tomamos o valor máximo da probabilidade *a posteriori*, dada a informação sobre o rendimento.

Esse é um exemplo do que se chama de modelo estático. Poderíamos considerar um modelo dinâmico, supondo-se que esse muda de período para período (de dia para dia ou de mês para mês etc.).

11.9. Exemplos Computacionais

Simulando Erros Padrões

Na seção 11.7 definimos o que seja o erro padrão de um estimador T de um parâmetro θ, baseado numa AAS de uma população rotulada pela v.a. X. Vimos, em particular, que o erro padrão da média amostral $\overline{X}$ é dado por (11.42) e esse pode ser estimado por (11.44), ou seja,

$$\widehat{\text{EP}}(\overline{X}) = \frac{S}{\sqrt{n}}.$$

O erro padrão de um estimador é fundamental para avaliarmos quão bom ele é. Simplesmente calcular T, ou saber que ele é não-viesado, não é suficiente: é necessário calcular sua variabilidade.

Mas, na maioria das situações, não podemos obter uma estimativa do erro padrão de um estimador. Considere, por exemplo, a mediana de uma amostra,

$$\text{md} = \text{med}(X_1, ..., X_n). \tag{11.49}$$

Pode não ser fácil calcular a Var(md) e, conseqüentemente, o erro padrão de md. Se admitirmos que a aproximação (11.18) é razoável, então teremos

$$\text{EP(md)} \approx \sigma \sqrt{\frac{\pi}{2n}}$$

e poderemos, novamente, estimar σ por S e obter

$$\widehat{\text{EP}}\text{(md)} \approx S \sqrt{\frac{\pi}{2n}}.$$

Mas, se tivermos amostras não muito grandes, a aproximação pode não ser adequada.

Felizmente, com o progresso de métodos computacionais usando intensivamente computadores cada vez mais rápidos e com capacidade cada vez maior de lidar com conjuntos grandes de dados, o cálculo de erros padrões, vieses etc., pode ser feito sem recorrer a uma teoria, que muitas vezes pode ser muito complicada ou simplesmente não existir.

Um desses métodos é chamado *bootstrap*, introduzido por B. Efrom, em 1979. Os livros de Efrom e Tibshirani (1993) e Davison e Hinkley (1997) são referências importantes para aqueles que quiserem se aprofundar no assunto.

A idéia básica do método *bootstrap* é re-amostrar o conjunto disponível de dados para estimar o parâmetro θ, com o fim de criar dados replicados. A partir dessas replicações, podemos avaliar a variabilidade de um estimador proposto para θ, sem recorrer a cálculos analíticos.

Vamos ilustrar o método com um exemplo.

Exemplo 11.19. Suponha que temos os dados amostrais $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ e queremos estimar a mediana populacional, Md, por meio da mediana amostral $\text{md}(\boldsymbol{x}) = \text{med}(x_1, ..., x_n)$.

Vamos escolher uma AAS (portanto, com reposição) de tamanho n dos dados. Tal amostra é chamada uma amostra *bootstrap* e denotada por $\boldsymbol{x}^* = (x_1^*, ..., x_n^*)$.

Por exemplo, suponha que $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5)$. Poderemos obter, por exemplo, $\boldsymbol{x}^* = (x_4, x_3, x_3, x_1, x_2)$.

Suponha, agora, que geremos B tais amostras independentes, denotadas $\boldsymbol{x}_1^*, ..., \boldsymbol{x}_B^*$. Para cada amostra *bootstrap*, geramos uma *réplica bootstrap* do estimador proposto, ou seja, de $\text{md}(\boldsymbol{x})$, obtendo-se

$$\text{md}(\boldsymbol{x}_1^*), \text{md}(\boldsymbol{x}_2^*), ..., \text{md}(\boldsymbol{x}_B^*). \tag{11.50}$$

Definimos o estimador *bootstrap* do erro padrão de $\text{md}(x)$ como

$$\widehat{\text{EP}}_B(\text{md}) = \left[\frac{\sum_{b=1}^{B}(\text{md}(\boldsymbol{x}_b^*) - \overline{\text{md}})^2}{B-1}\right]^{1/2}, \tag{11.51}$$

com

$$\overline{\mathrm{md}} = \frac{\sum_{b=1}^{B} \mathrm{md}(\boldsymbol{x}_b^*)}{B}. \qquad (11.52)$$

Ou seja, o estimador *bootstrap* do erro padrão da mediana amostral é o desvio padrão amostral do conjunto (11.50). Na Figura 11.5 temos representado o esquema do método.

Vamos ilustrar o método com um exemplo numérico simples. Suponha que $n = 5$ e a amostra é $\boldsymbol{x} = (2, 5, 3, 4, 6)$. Vamos considerar $B = 5$ amostras *bootstrap* de $\boldsymbol{x}$. Como gerar tais amostras? Primeiramente, geramos cinco números aleatórios $i_1, ..., i_5$ dentre os cinco números inteiros 1, 2, 3, 4, 5 e consideramos a amostra *bootstrap* $\boldsymbol{x}^* = (x_{i1}, ..., x_{i5})$. Repetimos esse procedimento cinco vezes. Podemos usar a Tabela VII para gerar esses NA, como já aprendemos. Considere, por exemplo, as cinco primeiras linhas e, começando do canto esquerdo, prossiga em cada linha até obter cinco dígitos entre 1 e 5, inclusive; note que pode haver repetições! Obtemos a Tabela 11.2.

Figura 11.5: Procedimento *bootstrap* para calcular o erro padrão da mediana amostral.

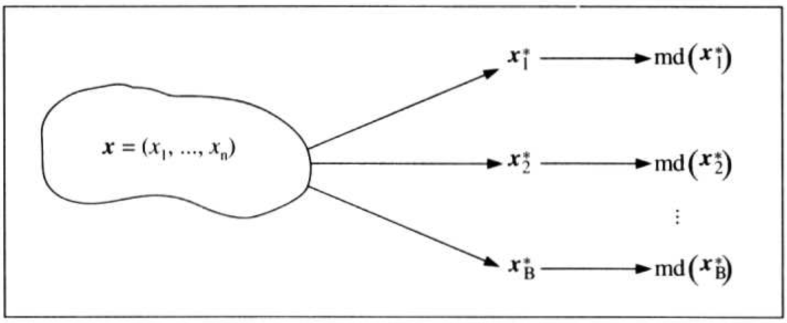


Tabela 11.2: Procedimento *bootstrap*.

NA	Amostra *bootstrap*	md($\boldsymbol{x}^*$)	$\bar{x}(\boldsymbol{x}^*)$
1,2,2,5,1	(2,5,5,6,2)	5,0	4,0
4,4,4,3,2	(4,4,4,3,5)	4,0	4,0
5,4,5,5,5	(6,4,6,6,6)	6,0	5,6
5,1,1,5,5	(6,2,2,6,6)	6,0	4,4
2,5,4,5,3	(5,6,4,6,3)	5,0	4,8

Por exemplo, obtidos os NA 1, 2, 2, 5, 1, teremos a amostra *bootstrap* $(x_1, x_2, x_2, x_5, x_1) = (2, 5, 5, 6, 2)$, para a qual a mediana amostral é 5. Segue-se que $\overline{\mathrm{md}} = 26/5 = 5{,}2$ e

$$\widehat{\mathrm{EP}}_B(\mathrm{md}) = \left[\frac{\sum_{b=1}^{5}(\mathrm{md}(\boldsymbol{x}_b^*) - 5{,}2)^2}{4}\right]^{1/2} = 0{,}837.$$

Se usarmos a aproximação (11.18), calculamos a variância da amostra original, obtendo-se $S^2 = 2{,}5$, donde $\widehat{\mathrm{EP}}(\mathrm{md}) \approx 0{,}886$. Levando-se em conta o tamanho da amostra, a discrepância entre os dois valores não é grande.

Exemplo 11.20. Na Tabela 11.2 calculamos, também, para cada amostra *bootstrap*, a média amostral, $\bar{x}$. Obtemos, usando (11.51),

$$\widehat{EP}_B(\bar{x}) = 0{,}669,$$

e usando a fórmula (11.44),

$$\widehat{EP}(\bar{x}) = \sqrt{2{,}5/5} = 0{,}707,$$

logo o valor obtido pelo método *bootstrap* está bastante próximo do valor calculado pela fórmula obtida de maneira analítica. Obviamente, em situações nas quais há uma fórmula disponível, não há necessidade de se usar *bootstrap*.

A questão que se apresenta é: qual deve ser o valor de B, ou seja, quantas amostras *bootstrap* devemos gerar para estimar erros padrões de estimadores? A experiência indica que um valor razoável é $B = 200$.

No caso geral de um estimador $\hat{\theta} = T(\boldsymbol{x})$, o algoritmo *bootstrap* para estimar o erro padrão de $\hat{\theta}$ é o seguinte:

[1] Selecione B amostras *bootstrap* independentes $\boldsymbol{x}_1^*, ..., \boldsymbol{x}_B^*$, cada uma consistindo de n valores selecionados com reposição de $\boldsymbol{x}$. Tome $B \approx 200$.

[2] Para cada amostra *bootstrap* $\boldsymbol{x}_b^*$ calcule a réplica *bootstrap*

$$\hat{\theta}^*(b) = T(\boldsymbol{x}_b^*), \quad b = 1, 2, ..., B.$$

[3] O erro padrão de $\hat{\theta}$ é estimado pelo desvio padrão das B réplicas:

$$\widehat{EP}_B = \left[\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^*(b) - \overline{\theta^*})^2\right]^{1/2}, \tag{11.53}$$

com

$$\overline{\theta^*} = \frac{\sum_{i=1}^{B} \hat{\theta}^*(b)}{B}. \tag{11.54}$$

No exemplo acima, notamos que um intervalo de confiança aproximado para a mediana populacional Md, com coeficiente de confiança 95%, seria

$$5{,}2 \pm (1{,}96)(0{,}837) =]3{,}56; 6{,}84[.$$

No exemplo dado, para efeito de ilustração do método *bootstrap*, tomamos uma amostra pequena ($n = 5$) e poucas amostras *bootstrap* ($B = 5$). Para amostras maiores e B na ordem de 200 deveremos fazer um pequeno programa, em alguma linguagem (como o Visual Basic, Pascal, Fortram, C etc.), que gere as amostras *bootstrap*, e calcular o estimador dado por (11.53). Isso implica, em particular, gerar, para cada amostra *bootstrap*, n números aleatórios. Como já vimos, não é prático usar uma tabela de NA nessa situação; devemos usar alguma rotina de computador.

11.10 Problemas e Complementos

22. Um pesquisador está em dúvida sobre duas possíveis estatísticas, t e t', para serem usadas como estimadores de um parâmetro θ. Assim, ele decidiu usar simulação para uma situação hipotética, procurando encontrar pistas que o ajudassem a decidir qual o melhor estimador. Partindo de uma população fictícia, onde $\theta = 10$, ele retirou 1.000 amostras de 20 elementos, e para cada amostra calculou o valor das estatísticas t e t'. Em seguida, construiu a distribuição de freqüências, segundo o quadro abaixo.

Classes	% de t	% de t'
5 ⊢ 7	10	5
7 ⊢ 9	20	30
9 ⊢ 11	40	35
11 ⊢ 13	20	25
13 ⊢ 15	10	5

(a) Verifique as propriedades de t e t' como estimadores de θ.
(b) Qual dos dois você adotaria? Por quê?

23. De experiências passadas, sabe-se que o desvio padrão da altura de crianças de 5ª série do 1º grau é 5 cm.
(a) Colhendo uma amostra de 36 dessas crianças, observou-se a média de 150 cm. Qual o intervalo de confiança de 95% para a média populacional?
(b) Que tamanho deve ter uma amostra para que o intervalo $150 \pm 0{,}98$ tenha 95% de confiança?

24. Um pesquisador está estudando a resistência de um determinado material sob determinadas condições. Ele sabe que essa variável é normalmente distribuída com desvio padrão de duas unidades.
(a) Utilizando os valores 4,9; 7,0; 8,1; 4,5; 5,6; 6,8; 7,2; 5,7; 6,2 unidades, obtidos de uma amostra de tamanho 9, determine o intervalo de confiança para a resistência média com um coeficiente de confiança $\gamma = 0{,}90$.
(b) Qual o tamanho da amostra necessário para que o erro cometido, ao estimarmos a resistência média, não seja superior a 0,01 unidade com probabilidade 0,90?
(c) Suponha que no item (a) não fosse conhecido o desvio padrão. Como você procederia para determinar o intervalo de confiança, e que suposições você faria para isso? Veja também o problema 44.

25. Estime o salário médio dos empregados de uma indústria têxtil, sabendo-se que uma amostra de 100 indivíduos apresentou os seguintes resultados:

Salário	Freqüência
150,00 ⊢ 250,00	8
250,00 ⊢ 350,00	22
350,00 ⊢ 450,00	38
450,00 ⊢ 550,00	28
550,00 ⊢ 650,00	2
650,00 ⊢ 750,00	2

Use $\gamma = 0{,}95$.

26. Suponha que as vendas de um produto satisfaçam ao modelo

$$V_t = \alpha + \beta t + a_t,$$

onde a_t é a variável aleatória satisfazendo as suposições da seção 11.4, e o tempo é dado em meses. Suponha que os valores das vendas nos 10 primeiros meses de 1979 sejam dados pelos valores da tabela abaixo. Obtenha as previsões para os meses de novembro e dezembro de 1979 e para julho e agosto de 1980.

t	1	2	3	4	5	6	7	8	9	10
y_t	5,0	6,7	6,0	8,7	6,2	8,6	11,0	11,9	10,6	10,8

27. Numa pesquisa de mercado para estudar a preferência da população de uma cidade em relação a um determinado produto, colheu-se uma amostra aleatória de 300 indivíduos, dos quais 180 preferiam esse produto.
   (a) Determine um intervalo de confiança para a proporção da população que prefere o produto em estudo; tome $\gamma = 0,90$.
   (b) Determine a probabilidade de que a estimativa pontual dessa proporção não difira do verdadeiro valor em mais de 0,001.
   (c) É possível obter uma estimativa pontual dessa proporção que não difira do valor verdadeiro em mais de 0,0005 com probabilidade 0,95? Caso contrário, determine o que deve ser feito.

28. Uma amostra de 10.000 itens de um lote de produção foi inspecionada, e o número de defeitos por item foi registrado na tabela abaixo.

Nº de defeitos	0	1	2	3	4
Quantidade de peças	6.000	3.200	600	150	50

   (a) Determine os limites de confiança para a proporção de itens defeituosos na população, com coeficiente de confiança de 98%. Use (11.40).
   (b) Mesmo problema, usando (11.41).

29. Antes de uma eleição em que existiam dois candidatos, A e B, foi feita uma pesquisa com 400 eleitores escolhidos ao acaso, e verificou-se que 208 deles pretendiam votar no candidato A. Construa um intervalo de confiança, com c.c. $\gamma = 0,95$, para a porcentagem de eleitores favoráveis ao candidato A na época das eleições.

30. Encontre o c.c. de um intervalo de confiança para p, se $n = 100$, $\hat{p} = 0,6$ e a amplitude do intervalo deve ser igual a 0,090.

31. Usando os resultados do problema 32 do Capítulo 10, mostre que o intervalo de confiança para a diferença das médias populacionais, com variâncias conhecidas, é dado por

$$IC(\mu_1 - \mu_2 : \gamma) = (\overline{X} - \overline{Y}) \pm z(\gamma)\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}.$$

32. Estão sendo estudados dois processos para conservar alimentos, cuja principal variável de interesse é o tempo de duração destes. No processo A, o tempo X de duração segue a distribuição $N(\mu_A, 100)$, e no processo B o tempo Y obedece à distribuição $N(\mu_B, 100)$.

Sorteiam-se duas amostras independentes: a de A, com 16 latas, apresentou tempo médio de duração igual a 50, e a de B, com 25 latas, duração média igual a 60.

(a) Construa um IC para μ_A e μ_B, separadamente.

(b) Para verificar se os dois processos podem ter o mesmo desempenho, decidiu-se construir um IC para a diferença $\mu_A - \mu_B$. Caso o zero pertença ao intervalo, pode-se concluir que existe evidência de igualdade dos processos. Qual seria sua resposta?

33. Numa pesquisa sobre a opinião dos moradores de duas cidades, A e B, com relação a um determinado projeto, obteve-se:

Cidade	A	B
Nº de entrevistados	400	600
Nº de favoráveis	180	350

Construa um IC para a diferença de proporções de opiniões nas duas cidades (Veja o problema 35 do Capítulo 10.)

34. Seja X uma v.a. com $E(X) = \mu$ e $Var(X) = \sigma^2$ finita. Então, para todo $k > 0$, a seguinte desigualdade (chamada desigualdade de Chebyshev) é válida:

$$P(|X - \mu| \geqslant k) \leqslant Var(X)/k^2. \tag{11.55}$$

Usando (11.55), prove que $\overline{X}$ é um estimador consistente para a média μ de uma população com variância σ^2.

35. Lei dos Grandes Números. Consideremos n provas de Bernoulli com $p = P$ (sucesso), e seja k o número de sucessos nas n provas. A Lei dos Grandes Números (LGN) afirma que, para n grande, a proporção de sucessos k/n estará próxima de $p = P$ (sucesso). Formalmente, para todo $\varepsilon > 0$,

$$P\left\{\left|\frac{k}{n} - p\right| \geqslant \varepsilon\right\} \leqslant \frac{p(1-p)}{n\varepsilon^2}. \tag{11.56}$$

Prove (11.56), usando (11.55).

36. A LGN pode ser usada de maneira útil na seguinte situação. Suponha que queiramos saber quantas repetições de um experimento de Bernoulli devemos realizar a fim de que k/n difira de p de menos de ε, com probabilidade maior ou igual a γ. Ou seja, queremos determinar n, tal que

$$P\left\{\left|\frac{k}{n} - p\right| < \varepsilon\right\} \geqslant \gamma.$$

De (11.56) temos

$$P\left\{\left|\frac{k}{n} - p\right| < \varepsilon\right\} \geqslant 1 - \frac{p(1-p)}{n\varepsilon^2};$$

logo, comparando, temos que n deve satisfazer

$$1 - \frac{p(1-p)}{n\varepsilon^2} = \gamma \Rightarrow n = \frac{p(1-p)}{\delta\varepsilon^2}, \text{ onde } \delta = 1 - \gamma.$$

Como não conhecemos p, usando o fato de que $p(1 - p) \leqslant 1/4$; logo, basta tomar n tal que $n = 1/4\delta\varepsilon^2$.

Usando esse resultado, resolva este problema: suponha que a proporção de fumantes de uma população é p, desconhecida. Queremos determinar p com um erro de, no máximo, 0,05. Qual deve ser o tamanho da amostra n, a ser escolhida com reposição, se $\gamma = 0{,}95$?

37. Se a distribuição de X depende de mais de um parâmetro, digamos θ_1 e θ_2, então $L(\theta_1, \theta_2; X_1, ..., X_n)$, e para maximizar L basta derivar L em relação a θ_1 e θ_2 (em algumas situações, derivar L não conduz ao EMV; veja o problema 43). Considere, então, $X \sim N(\mu, \sigma^2)$. Determine os EMV de μ e σ^2, considerando $\partial\ell/\partial\mu = 0$ e $\partial\ell/\partial\sigma^2 = 0$, onde $\ell = \log L$.

38. Estimação numa distribuição uniforme. Suponha que X tenha uma distribuição uniforme no intervalo $(0, \theta)$, onde θ é desconhecido. Uma amostra de n observações $X_1, ..., X_n$ é escolhida. Sabemos que $E(X) = E(X_i) = \theta/2$, para todo i, e $\text{Var}(X) = \text{Var}(X_i) = \theta^2/12$, para todo i. Logo, se calcularmos a média amostral $\overline{X}$, essa deve estar próxima de $\theta/2$ e podemos estimar θ por $T_1 = 2\overline{X}$.
   (a) Calcule $E(T_1)$.
   (b) Calcule $\text{EQM}(T_1) = E(T_1 - \theta)^2$.
   (c) T_1 é consistente? Por quê?

39. Continuação do problema 38. Outra maneira de estimar θ na uniforme é a seguinte. Considere $M = \max(X_1, ..., X_n) = x_{(n)}$, ou seja, o maior valor da amostra. Para qualquer valor de θ, $M < \theta$ e M se aproxima de θ quando n aumenta. Tome M como estimador de θ, o que é bastante razoável. Na realidade, veremos, no problema 43, que $M = \hat{\theta}_{MV}$. Vimos no problema 38 do Capítulo 10, que a densidade de M é dada por

$$f_M(x) = \begin{cases} \dfrac{n}{\theta^n} x^{n-1}, & \text{se } 0 \leq x \leq \theta \\ 0, & \text{caso contrário.} \end{cases} \tag{11.57}$$

   (a) Mostre que $E(M) = \theta \dfrac{n}{n+1}$, logo M é viesado. Calcule o viés $V_M(\theta)$ de M e mostre que esse viés tende a zero, quando $n \to \infty$.
   (b) Considere o estimador $T_2 = \dfrac{n+1}{n} M$; segue-se que T_2 é não-viesado para θ, ou seja, $E(T_2) = \theta$. Calcule o erro quadrático médio de T_2, $\text{EQM}(T_2) = E(T_2 - \theta)^2$.
   (c) T_2 é consistente? Por quê?

40. Para os problemas 38 e 39, mostre que $\text{Var}(T_2) = [3/(n + 2)]\ \text{Var}(T_1)$. Tome $n = 1, 2, 10, 50, 100$ e verifique qual a relação entre as duas variâncias. Verifique que, para n grande, $T_2 = [(n + 1)/n]M$ é um estimador muito melhor do que $T_1 = 2\overline{X}$. Como $T_2 = (1 + 1/n)M$, vemos que, para n grande, $T_2 \approx M$. Portanto, para tamanhos de amostras grandes, o EMV é melhor do que $2\overline{X}$.

41. Considere as situações dos problemas 38, 39 e 40. Suponha que n seja suficientemente grande para que o Teorema Limite Central se aplique e se possa aproximar a distribuição de $\bar{X}$ por uma normal. Obtenha um intervalo de confiança para θ, com c.c. = 90%:
   (a) usando T_1 como estimador de θ;
   (b) usando T_2 como estimador de θ;
   (c) usando M como estimador de θ;
   (d) qual a relação entre os intervalos de (b) e (c) se n for muito grande?
   [Sugestão: raciocine como no caso de $\hat{p}$; nesse caso, vimos que $\hat{p} \sim N(p,\ p(1-p)/n)$ e estimamos a variância de $\hat{p}$ por $\hat{p}(1-\hat{p})/n$.]

42. Foram gerados 1.000 valores de uma distribuição uniforme no intervalo $(0, 5)$, ou seja, $\theta = 5$. As seguintes estatísticas foram obtidas:

$$x_{(1)} = min(X_1, ..., X_{1000}) = 0{,}01132, \quad x_{(1000)} = M = max(X_1, ..., X_{1000}) = 4{,}992;$$

$$q_1 = 1{,}315, \quad q_2 = 2{,}572, \quad q_3 = 3{,}829, \quad \bar{x} = 2{,}547.$$

   Calcule T_1, T_2 e aplique o resultado do problema 41 para obter um intervalo de confiança para θ, com c.c. = 90%.

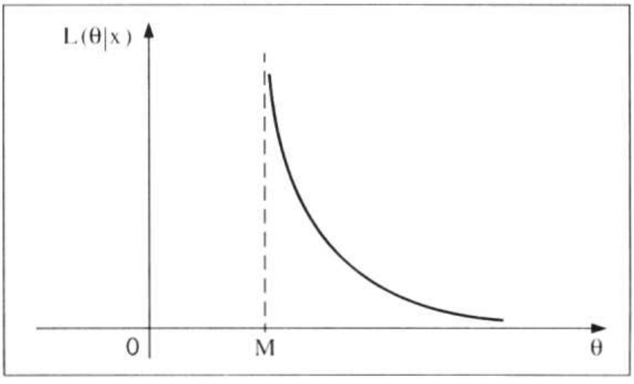
43. EMV na uniforme. Como

$$f(x) = \begin{cases} 1/\theta, & \text{se } 0 \leq x \leq \theta, \\ 0, & \text{caso contrário,} \end{cases}$$

a densidade conjunta da amostra é

$$f(x_1, ..., x_n; \theta) = \begin{cases} 1/\theta^n, & \text{se } 0 \leq x_i \leq \theta, i = 1, ..., n \\ 0, & \text{caso contrário.} \end{cases}$$

Segue-se que $\ell(\theta | x_1, ..., x_n) = -n \log \theta$ e derivando e igualando a zero obteremos $-n/\theta = 0$, ou seja, o EMV de θ seria ∞! Evidentemente, essa não é a resposta. Na realidade, não podemos simplesmente derivar a verossimilhança (ou o logaritmo dela) para obter o máximo, pois temos as restrições $0 \leq x_i \leq \theta$, para todo i. Façamos o seguinte. Considere o gráfico da densidade conjunta, ou da verossimilhança, como função de θ. Como devemos ter $0 \leq x_i \leq \theta$, para todo i, o máximo M dos x_i deve ser tal que $0 \leq M \leq \theta$, ou seja, obtemos o gráfico abaixo.



Ou seja, $L(\theta | x_1, ..., x_n) = 0$, para $\theta \leqslant M$; logo, o máximo da verossimilhança é obtido para $\theta = M$ e portanto $\hat{\theta}_{MV} = M$.

Esse exemplo mostra que nem sempre obteremos o EMV derivando-se a verossimilhança e igualando-a a zero.

44. Suponha que $X \sim N(\mu, \sigma^2)$, μ e σ^2 desconhecidos. Uma amostra de tamanho $n = 600$ forneceu $\overline{X} = 10{,}3$ e $S^2 = 1{,}96$. Supondo que a v.a. $Z = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ seja aproximadamente normal, obtenha um IC para μ, com c.c. $\gamma = 0{,}95$ (se n for pequeno, Z não é aproximadamente normal; ver Capítulo 12).

45. Para estimar a média μ desconhecida de uma população, foram propostos dois estimadores não-viesados independentes, $\hat{\mu}_1$ e $\hat{\mu}_2$, de tal sorte que $\mathrm{Var}(\hat{\mu}_1) = \mathrm{Var}(\hat{\mu}_2)/3$. Considere os seguintes estimadores ponderados de μ:

    (a) $T_1 = (\hat{\mu}_1 + \hat{\mu}_2)/2$;
    (b) $T_2 = (4\hat{\mu}_1 + \hat{\mu}_2)/5$;
    (c) $T_3 = \hat{\mu}_1$.

    (i) Quais estimadores são não-viesados?
    (ii) Dispor esses estimadores em ordem crescente de eficiência.

46. Obtenha o estimador de λ na Poisson, pelo método dos momentos.

47. Considere o CD-Notas e retire uma amostra com reposição de tamanho $n = 10$. Determine o erro padrão estimado pelo método *bootstrap* das estatísticas (use $B = 15$, por exemplo):

    (a) md = mediana da amostra;
    (b) dm = desvio médio da amostra.
    (c) dam = desvio absoluto mediano.