2 Estimação Paramétrica Pontual

2.1 Introdução

2.1.1 Aspectos Gerais

Situamo-nos no âmbito da inferência paramétrica:

- População X tem distribuição na família $\mathcal{F} = \{f(\cdot \mid \theta) : \theta \in \Theta\}$
- Espaço do paramétrico $\Theta \subset \mathbb{R}^k$
- Define-se processo de amostragem; em geral $(X_1, \ldots, X_n)$ amostra casual de tamanho n proveniente de X
- Espaço amostral $\mathcal{X}$

Problema:

- estimar θ pontualmente, ou seja, escolher uma aplicação $\mathbf{x} \in \mathcal{X} \mapsto T(\mathbf{x}) \in \Theta$ que a cada amostra observada faz corresponder um valor para o parâmetro
- A aplicação T é uma estatística a que se dá o nome de <u>estimador</u> de θ; cada ponto $T(\mathbf{x})$ é designado por <u>estimativa</u>
- Podemos por vezes estar interessados em estimar uma função de θ, $\tau(\theta)$

Exemplo 2.1 *Admite-se que X, o número de sinistros originados anualmente por uma apólice de seguro automóvel, tem distribuição de Poisson de parâmetro (desconhecido) $\lambda > 0$.*

Podemos estar interessados em estimar λ. Um estimador natural é $T(X_1, \ldots, X_n) = \bar{X}$; a estimativa de λ associada é $T(x_1, \ldots, x_n) = \bar{x}$.

Em vez de focarmos o nosso interesse no número médio de sinistros por apólice, podemos estar interessados em $\tau(\lambda) = e^{-\lambda}$ que traduz a probabilidade de não se verificar nenhum sinistro: $P(X = 0 \mid \lambda) = e^{-\lambda}$. ■

A definição de estimador é extremamente geral:

- o que se entende por um "bom" estimador? — critérios de optimalidade
- métodos de construção de estimadores

2.1.2 Suficiência

- Conceito cujo interesse extravasa o problema da estimação pontual
- Ideia intuitiva: uma estatística é suficiente para o parâmetro θ se retira da amostra aleatória $X_1, \ldots, X_n$ toda a informação relevante que esta contém sobre o parâmetro desconhecido θ

Exemplo 2.1 *(Continuação) Suponhamos que se observa o número de sinistros associados a* $n = 10$ *apólices de seguro automóvel,* $x_1, \ldots, x_{10}$*, e que se anota que* $\sum_{i=1}^{10} x_i = 14$*. Devido a um erro no sistema informático da seguradora, perde-se a amostra* $x_1, \ldots, x_{10}$*. Os responsáveis pelo erro pretendem produzir uma amostra* $x_1^\star, \ldots, x_{10}^\star$ *tal que* $\sum_{i=1}^{10} x_i^\star = 14$ *e que seja estatisticamente indistinguível da originalmente observada. Isso é possível?*

Seja $T = \sum_{i=1}^{n} X_i$ e $n = 10$. Qual a distribuição de $X_1, \ldots, X_n \mid T = 14$?

$$
\begin{aligned}
P(X_1 = x_1, \ldots, X_n = x_n \mid T = 14) &= \frac{P(X_1 = x_1, \ldots, X_n = x_n, T = 14)}{P(T = 14)} \\
&= \begin{cases} \dfrac{P(X_1 = x_1, \ldots, X_n = x_n)}{P(T = 14)} & \text{se } \sum_{i=1}^{n} x_i = 14 \\ 0 & \text{se } \sum_{i=1}^{n} x_i \neq 14 \end{cases} \\
&= \frac{e^{-n\lambda} \lambda^{\sum x_i} / \prod x_i!}{e^{-n\lambda} (n\lambda)^{14} / 14!} \quad \text{se } \sum_{i=1}^{n} x_i = 14 \\
&= \frac{14!}{\prod_{i=1}^{n} x_i!} \, \frac{1}{n} \cdots \frac{1}{n} \quad \text{se } \sum_{i=1}^{n} x_i = 14
\end{aligned}
$$

Ou seja, $X_1, \ldots, X_n \mid T = 14$ segue uma distribuição multinomial com vector de probabilidade $\psi = (1/n, \ldots, 1/n)$. Esta distribuição não depende de λ! Sabendo que se observaram 14 sinistros, não é necessário conhecer λ para gerar a amostra $x_1^\star, \ldots, x_n^\star$.

Diz-se neste caso que T é suficiente para λ: toda a informação relevante sobre λ disponível em $X_1 \ldots, X_n$ está contida em T.

Definição 2.1 Estatística suficiente: *Seja* $(X_1, \ldots, X_n)$ *uma amostra casual proveniente de uma população* $\mathcal{F} = \{f(\cdot \mid \theta) : \theta \in \Theta\}$. *Diz-se que a estatística* T *é suficiente para* $\mathcal{F}$ *(ou, simplesmente, para* θ*) se a distribuição de* $(X_1, \ldots, X_n)$ *condicionada por* $T = t$ *não depende de* θ, $\forall \theta \in \Theta$ *e* $\forall t$. ■

Exemplo 2.2 *Seja* $(X_1, \ldots, X_n)$ *uma amostra casual de dimensão* $n > 3$ *proveniente de uma população* $B(1, \theta)$.

Note-se que

$$f(x_1, \ldots, x_n \mid \theta) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}, \quad x_i = 0, 1, \; i = 1, \ldots, n \; .$$

Como $T = \sum_{i=1}^n X_i \sim B(n, \theta)$, *tem-se que, se* $\sum_{i=1}^n x_i = t$

$$f(x_1, \ldots, x_n \mid t) = 1/\binom{n}{t}$$

que não depende de θ. *Toda a informação relevante sobre* θ *contida em* $(X_1, \ldots, X_n)$ *está condensada em* T: T *é suficiente para* θ.

Qualquer função biunívoca de T *é também suficiente para* θ, *por exemplo* $\bar{X}$.

$S = X_1 + X_2$ *não é suficiente para* θ: *se* $S = s$,

$$f(x_1, \ldots, x_n \mid s) = \theta^{\sum_{i=3}^n x_i}(1-\theta)^{n-2-\sum_{i=3}^n x_i}/\binom{2}{s}$$

■

A definição de suficiência é útil para verificar se determinada estatística, cuja distribuição por amostragem conhecemos, é ou não suficiente para θ. A sua utilidade na pesquisa de estatísticas suficientes é limitada.

Teorema 2.1 Critério da Factorização: *A estatística T é suficiente para θ se e só se existem duas funções não negativas $g(\cdot)$ e $h(\cdot)$ tais que para todo o $x_1, \ldots, x_n$*

$$f(x_1, \ldots, x_n \mid \theta) = g[T(x_1, \ldots, x_n), \theta] \times h(x_1, \ldots, x_n) \ .$$

■

Note-se que $g(\cdot)$ depende de $(x_1, \ldots, x_n)$ unicamente através de $T(x_1, \ldots, x_n)$, e de θ, enquanto que $h(\cdot)$ depende unicamente de $(x_1, \ldots, x_n)$.

Exemplo 2.3 *Seja $(X_1, \ldots, X_n)$ uma amostra casual de dimensão n proveniente de uma população $Po(\lambda)$. Então,*

$$\begin{aligned} f(x_1, \ldots, x_n \mid \lambda) &= \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \\ &= e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i} \times \left(\prod_{i=1}^{n} x_i! \right)^{-1} \end{aligned}$$

Pelo Critério da Factorização, $T = \sum_{i=1}^{n} X_i$ é suficiente para λ. ■

Exemplo 2.4 $X_1, \ldots, X_n$ *amostra casual proveniente de população* $N(\mu, \sigma^2)$. *Então,*

$$\begin{aligned} f(x_1, \ldots, x_n \mid \mu, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x_i - \mu)^2\right] \\ &= (2\pi\sigma^2)^{-n/2} \ \exp[-n\mu^2/(2\sigma^2)] \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 - 2n\mu \sum_{i=1}^{n} x_i\right)\right] \end{aligned}$$

logo $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ *é suficiente para* (μ, σ^2), *e consequentemente* $(\bar{X}, S'^2)$ *também é suficiente para* (μ, σ^2).

Exemplo 2.5 $X_1, \ldots, X_n$ *amostra casual proveniente de população* $U(0, \theta)$, $\theta > 0$.

$$\begin{aligned} f(x_1, \ldots, x_n \mid \theta) &= \prod_{i=1}^{n} \frac{1}{\theta} \quad se\ x_i \in (0, \theta),\ i = 1, \ldots, n \\ &= \theta^{-n} \prod_{i=1}^{n} I_{(0,\theta)}(x_i) \\ &= \theta^{-n} I_{(-\infty,\theta)}(x_{(n)}) \times I_{(0,+\infty)}(x_{(1)}) \end{aligned}$$

concluindo-se que $X_{(n)}$ *é suficiente para* θ. ■

2.1.3 Verosimilhança e Informação

Consideremos modelos uniparamétricos $\mathcal{F} = \{f(\cdot \mid \theta) : \theta \in \Theta\}$, onde $\Theta \subset \mathbb{R}$. Seja $X_1, \ldots, X_n$ uma amostra casual de tamanho n proveniente de $\mathcal{F}$.

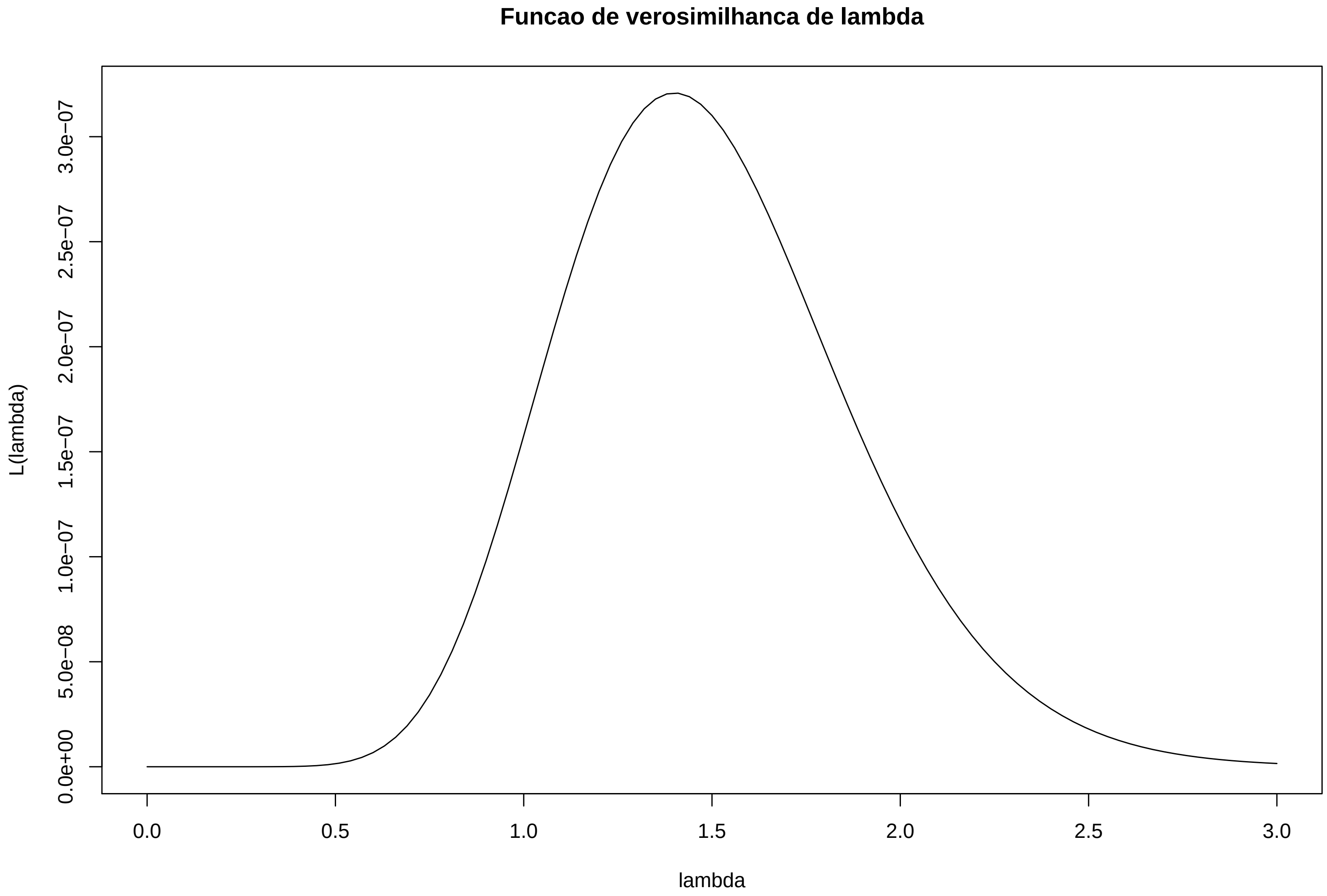
Definição 2.2 Verosimilhança: *Observada uma amostra $(x_1, \ldots, x_n)$, a correspondente função de verosimilhança de θ é dada por*

$$\begin{aligned} L(\theta \mid x_1, \ldots, x_n) &= f(x_1, \ldots, x_n \mid \theta) \\ &= \prod_{i=1}^{n} f(x_i \mid \theta) \ . \end{aligned}$$

■

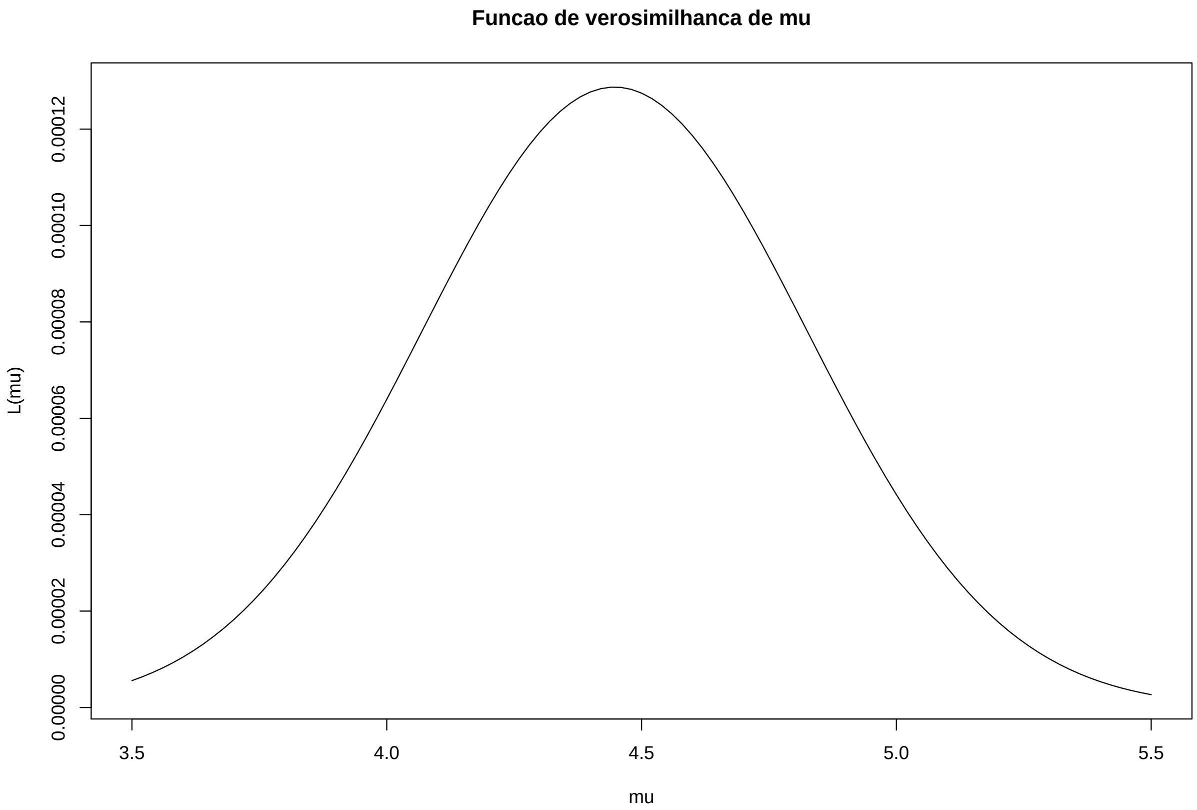
Como $x_1, \ldots, x_n$ está fixo, é comum escrever-se $L(\theta)$.

Exemplo 2.6 *Suponha-se que $X \mid \lambda \sim Po(\lambda)$ representa o número de sinistros originados anualmente por uma apólice de seguro automóvel. Observa-se uma amostra casual de tamanho $n = 10$ tal que $\sum_{i=1}^{n} x_i = 14$ e $\prod x_i! = 288$. Tem-se $L(\lambda) = \exp(-10\lambda)\lambda^{14}/288$.*


Funcao de verosimilhanca de lambda
L(lambda)
lambda
0.0e+00
5.0e−08
1.0e−07
1.5e−07
2.0e−07
2.5e−07
3.0e−07
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Exemplo 2.7 *Suponha-se que* $X \mid \mu \sim N(\mu, 1)$ *e que se observou a amostra* $(3.38, 5.93, 4.75, 3.77, 4.18, 5.29, 3.83)$. *Como* $\bar{x} = 4.45$ *e* $\sum x_i^2 = 143.49$, *segue-se que*

$$L(\mu) = (2\pi)^{-7/2} \exp\left[-\frac{1}{2}\left(143.49 - 2 \times 7 \times 4.45\mu + 7\mu^2\right)\right]$$



Observação importante: A verosimilhança não é uma função (densidade de) probabilidade e portanto não tem uma escala natural associada. Está definida portanto a menos de uma constante multiplicativa:

$$L(\theta \mid x_1, \ldots, x_n) \propto f(x_1, \ldots, x_n \mid \theta)$$

constante essa que pode depender de $x_1, \ldots, x_n$ mas que não depende de θ. O que tem significado são razões de verosimilhança

$$\frac{L(\theta' \mid x_1, \ldots, x_n)}{L(\theta^\star \mid x_1, \ldots, x_n)}$$

medindo quão verosímel é θ' comparativamente com $\theta^\star$ à luz dos dados observados $x_1, \ldots, x_n$.

Condições de regularidade

C1—Θ é um intervalo aberto de $\mathbb{R}$.

C2—O conjunto $\{x : f(x \mid \theta) > 0\}$, i.e., o suporte de $f(\cdot \mid \theta)$, não depende de θ.

C3—A função $f(x \mid \theta)$, $x \in \mathcal{X}$, $\theta \in \Theta$, é diferenciável em ordem a θ para todo o x

C4—Tem-se que $0 < E\left[\partial \ln f(X \mid \theta)/\partial\theta\right]^2 < +\infty$ para todo o θ

C5—A permutação de $\frac{\partial}{\partial\theta}$ and $\int dx$ é válida.

Observações:

- C2 exclui modelos como $\mathrm{U}(0, \theta)$: $\theta > 0$
- C4 garante que a v.a. $S = \partial \ln f(X \mid \theta)/\partial\theta$ tem segundo momento finito.

Definição 2.3 Função Score: *Observada a amostra* $x_1 \ldots, x_n$*, a função score destina-se a medir a variação relativa da função de verosimilhança em função de* θ*:*

$$S(\theta \mid x_1, \ldots, x_n) = \frac{L'(\theta \mid x_1, \ldots, x_n)}{L(\theta \mid x_1, \ldots, x_n)} = \frac{\partial \ln L(\theta \mid x_1, \ldots, x_n)}{\partial \theta}.$$

■

Nota: Para uma amostra casual, tem-se $S(\theta \mid x_1, \ldots, x_n) = \sum_{i=1}^{n} S(\theta \mid x_i)$ onde $S(\theta \mid x_i)$ é a função score associada à i-ésima observação.

Teorema 2.2 *Verificando-se as condições de regularidade, tem-se que para todo o* $\theta \in \Theta$

$$E[S(\theta \mid X_1, \ldots, X_n)] = 0 \; .$$

■

Definição 2.4 Medida de informação de Fisher: *A quantidade de informação de Fisher sobre θ contida na observação de $X_1, \ldots, X_n$ é definida por*

$$\mathcal{I}_{(X_1,\ldots,X_n)}(\theta) = E\left\{[S(\theta \mid X_1, \ldots, X_n)]^2\right\}$$

■

Propriedades:

- Verificando-se as condições de regularidade, $\mathcal{I}_{(X_1,\ldots,X_n)}(\theta) = \mathrm{Var}[S(\theta \mid X_1, \ldots, X_n)]$
- No caso de uma amostra casual, $\mathcal{I}_{(X_1,\ldots,X_n)}(\theta) = n\mathcal{I}_{X_i}(\theta)$
- Fórmula útil para o cálculo de $\mathcal{I}$: verificando-se as condições de regularidade,

$$\mathcal{I}_{(X_1,\ldots,X_n)}(\theta) = -E\left[\frac{\partial^2 \ln f(X_1, \ldots, X_n \mid \theta)}{\partial\theta^2}\right]$$

Teorema 2.3 *Sendo $T(X_1, \ldots, X_n)$ uma estatística, tem-se que*

$$\mathcal{I}_{(X_1,\ldots,X_n)}(\theta) \geq \mathcal{I}_T(\theta)$$

verificando-se a igualdade se e só se T for suficiente para θ. ■

Exemplo 2.8 *Seja $X_1, \ldots, X_n$ uma amostra aleatória proveniente de uma população $Po(\lambda)$, $\lambda > 0$. Então,*

$$S(\lambda \mid x_1, \ldots, x_n) = \frac{n}{\lambda}(\bar{x} - \lambda)$$

logo,

$$\mathcal{I}_{(X_1,\ldots,X_n)}(\lambda) = \left(\frac{n}{\lambda}\right)^2 Var(\bar{X}) = \frac{n}{\lambda}.$$

É fácil ver que $\mathcal{I}_{(X_1,\ldots,X_n)}(\lambda) = n\mathcal{I}_{X_i}(\lambda)$.

Além disso,

$$\frac{\partial}{\partial\lambda} S(\lambda \mid x_1, \ldots, x_n) = -\frac{\sum x_i}{\lambda^2}$$

pelo que

$$E\left[-\frac{\partial^2 \ln f(X_1, \ldots, X_n \mid \lambda)}{\partial\lambda^2}\right] = \frac{n}{\lambda} = \mathcal{I}_{(X_1,\ldots,X_n)}(\lambda)$$

■

2.2 Critérios de optimalidade

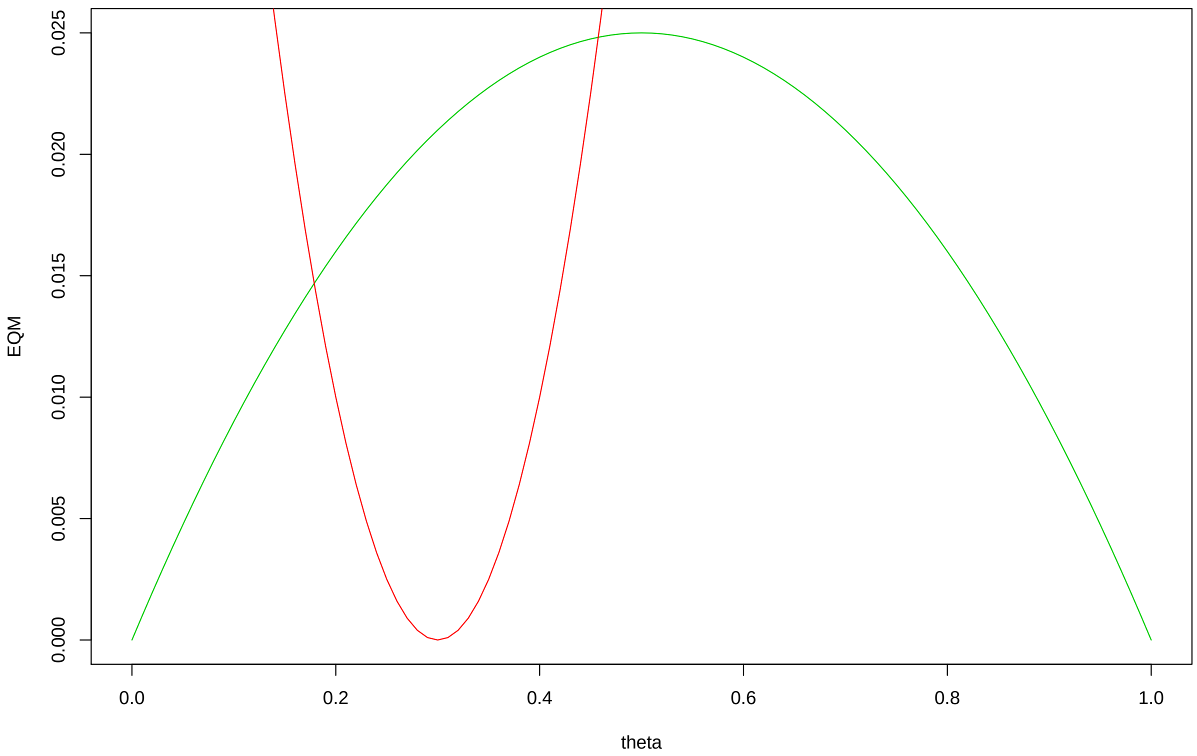
- A avaliação da qualidade de um estimador faz-se considerando a população de estimativas que gera, ou seja, a sua distribuição por amostragem. Avalia-se o estimador e não a estimativa.
- A existência de um estimador que seja 'ideal' é facilmente rebatida:

Exemplo 2.9 *Considere-se uma população $B(1,\theta)$ e uma amostra casual $X_1,\ldots,X_n$ dela extraída. É natural considerar que um estimador T é 'melhor' que um estimador S como estimadores de θ se*

$$E[(T-\theta)^2] \leq E[(S-\theta)^2] \quad \forall\theta$$

ou seja, se o erro quadrático médio de T nunca é superior ao de S.

Seja $T=\bar{X}$ e considere-se $S=0.3$. Claramente, $EQM_\theta(T)=\theta(1-\theta)/n$ e $EQM_\theta(S)=(\theta-0.3)^2$. ■


EQM
0.000
0.005
0.010
0.015
0.020
0.025
0.0
0.2
0.4
0.6
0.8
1.0
theta

2.2.1 Centragem

Definição 2.5 Estimador centrado: *Um estimador T diz-se um estimador centrado de $\tau(\theta)$ sse*

$$E[T] = \tau(\theta) \quad \forall\theta \in \Theta$$

■

Observações

- significa na prática que se se utilizar T para estimar $\tau(\theta)$ um número muito grande de vezes, então a média das estimativas obtidas estará próxima de $\tau(\theta)$ qualquer que seja θ
- $b_\theta(T) = E[T] - \tau(\theta)$ - enviesamento ou viés. Estimador não centrado diz-se enviesado.
- Nem sempre existem: observa-se $X \sim \mathrm{B}(1, \theta)$ e pretende-se um estimador centrado de θ^2, $T(X)$. Então, deve ter-se, $\forall\theta$

  $$\theta^2 = T(1)\theta + T(0)(1-\theta) \Leftrightarrow \theta^2 = [T(1) - T(0)]\theta + T(0)$$

  o que é impossível.
- A média amostral é sempre um estimador centrado da média da população, caso esta exista
- A variância amostral corrigida é sempre um estimador centrado da variância da população, caso esta exista

2.2.2 Eficiência

Definição 2.6 *Sejam T e $T^\star$ estimadores centrados de $\tau(\theta)$. Dizemos que T é mais eficiente que $T^\star$ na estimação de $\tau(\theta)$ se*

$$Var(T) \leq Var(T^\star) \quad \forall \theta \in \Theta$$

O estimador de $\tau(\theta)$ centrado T é o mais eficiente se qualquer que seja $T^\star$, estimador centrado de $\tau(\theta)$, se tem que T é mais eficiente que $T^\star$. ■

Teorema 2.4 Desigualdade de Fréchet-Cramer-Rao: *Considere-se uma população verificando as condições de regularidade, e seja $\tau(\theta)$ uma função real e diferenciável de θ. Seja ainda T um estimador centrado de $\tau(\theta)$ com variância finita. Se, $\forall \theta \in \Theta$*

$$E\left[\left|T(X_1,\ldots,X_n)\frac{\partial \ln f(X_1,\ldots,X_n \mid \theta)}{\partial\theta}\right|\right] < +\infty$$

$$E\left[T(X_1,\ldots,X_n)\frac{\partial \ln f(X_1,\ldots,X_n \mid \theta)}{\partial\theta}\right] = \frac{\partial}{\partial\theta}E[T(X_1,\ldots,X_n)]$$

caso em que T se diz regular, então

$$Var(T) \geq \frac{[\tau'(\theta)]^2}{\mathcal{I}_{(X_1,\ldots,X_n)}(\theta)} \ .$$

■

Observações:

- No caso em que $\tau(\theta) = \theta$ e $X_1, \ldots, X_n$ é amostra casual de população X, tem-se para qualquer estimador (regular) centrado de θ, T,

$$\mathrm{Var}(T) \geq [n\mathcal{I}_X(\theta)]^{-1}$$

- O limite inferior de FCR só é válido caso se verifiquem as condições de regularidade
- Mesmo verificando-se as condições de regularidade, não existem garantias de que exista um estimador centrado de $\tau(\theta)$ cuja variância atinja o limite inferior de FCR
- ao quociente entre o limite inferior de FCR e a variância de um estimador centrado de $\tau(\theta)$ dá-se o nome de eficiência (absoluta):

$$e_\theta(T) = \frac{[\tau'(\theta)]^2/\mathcal{I}_{(X_1,\ldots,X_n)}(\theta)}{\mathrm{Var}(T)}$$

  Se as condições de regularidade forem satisfeitas, tem-se que $0 \leq e_\theta(T) \leq 1$. Se $e_\theta(T) = 1$, T é necessariamente o estimador mais eficiente na estimação de $\tau(\theta)$.

- Fala-se ainda de eficiência assintótica, $e^\star_\theta(T) = \lim_{n\to+\infty} e_\theta(T)$, e em estimadores assintoticamente mais eficientes.

Exemplo 2.10 *Seja $X_1, \ldots, X_n$ uma amostra casual de tamanho n proveniente de uma população $Po(\lambda)$. Vimos que $\mathcal{I}_X(\lambda) = 1/\lambda$, logo qualquer estimador centrado de λ, T, verifica*

$$Var[T] \geq \lambda/n$$

Consequentemente, $\bar{X}$ é o estimador mais eficiente de λ.

Qualquer estimador centrado de $\tau(\lambda) = e^{-\lambda} = P(X = 0)$, S, verifica

$$Var(S) \geq \lambda e^{-2\lambda}/n \ .$$

Exemplo 2.11 *Seja $X \sim U(0, \theta)$, $\theta > 0$ e considere-se*

$$T = \frac{n+1}{n} X_{(n)} \ .$$

Verifica-se que T é centrado para θ e que $Var(T) = \theta^2/[n(n+1)]$. Além disso, $\mathcal{I}_{(X_1,\ldots,X_n)}(\theta) = n\theta^{-2}$ (calculada como se as condições de regularidade estivessem verificadas), pelo que

$$Var(T) < LIFCR.$$

Não estão satisfeitas as condições de regularidade. ■

Em que condições existem estimadores mais eficientes para $\tau(\theta)$? Corolário do Teorema de Frechét-Cramer-Rao:

Teorema 2.5 *Seja $T = T(X_1, \ldots, X_n)$ um estimador regular não enviesado de $\tau(\theta)$. T é mais eficiente se e só se existir $\beta(\theta)$ tal que*

$$S(\theta \mid x_1, \ldots, x_n) = \beta(\theta)[T(x_1, \ldots, x_n) - \tau(\theta)] \; .$$

■

Observações:

- se existir um estimador mais eficiente para $\tau(\theta)$, este será necessariamente suficiente
- caso não exista um estimador mais eficiente para $\tau(\theta)$, o conceito de eficiência (absoluta) deixa de fazer sentido
- eficiência relativa: T e $T^\star$ estimadores centrados, a eficiência de $T^\star$ relativamente a T é definida por

$$e_\theta(T^\star, T) = \frac{\mathrm{Var}(T)}{\mathrm{Var}(T^\star)}$$

Exemplo 2.12 *Se* $X \sim B(1, \theta)$*,*

$$S(\theta \mid x_1, \ldots, x_n) = \frac{n}{\theta(1-\theta)}(\bar{x} - \theta)$$

logo, $\bar{X}$ *é o estimador mais eficiente para* θ*.* ■

Exemplo 2.13 *Considere-se o modelo Cauchy de parâmetro de localização* θ *e parâmetro de escala 1, ou seja,* $f(x \mid \theta) = \{\pi[1 + (x-\theta)^2]\}^{-1}$*. Tem-se*

$$S(\theta \mid x_1, \ldots, x_n) = 2\sum_{i=1}^{n} \frac{x_i - \theta}{1 + (x_i - \theta)^2} \ .$$

Não existe estimador mais eficiente de θ*.* ■

2.2.3 Erro quadrático médio

Como comparar estimadores enviesados em termos da dispersão em torno de $\tau(\theta)$?

Definição 2.7 Erro quadrático médio *O erro quadrático médio de um estimador T de $\tau(\theta)$ é definido por*

$$EQM_\theta(T) = E[(T - \tau(\theta))^2] \ .$$

■

Observações:

- T será superior a $T^\star$ em erro quadrático médio na estimação de $\tau(\theta)$ se $\mathrm{EQM}_\theta(T) \leq \mathrm{EQM}_\theta(T^\star)$, $\forall \theta \in \Theta$
- Pela desigualdade de Markov, $P(|T - \tau(\theta)| > \varepsilon) \leq \mathrm{EMQ}_\theta(T)/\varepsilon^2$
- Tem-se que
$$\mathrm{EQM}_\theta(T) = \mathrm{Var}(T) + [b_\theta(T)]^2$$
- População normal: $E[S^{'2}] = \sigma^2$ e $E[S^2] = (n-1)\sigma^2/n$, no entanto,
$$\frac{2\sigma^4}{n-1} = \mathrm{Var}(S^{'2}) = \mathrm{EQM}_{\sigma^2}(S^{'2}) > \mathrm{EQM}_{\sigma^2}(S^2) = \frac{2\sigma^4}{n} - \frac{\sigma^4}{n^2}$$

2.2.4 Consistência

Comportamento probabilístico do estimador à medida que $n \to +\infty$: seja $T_n = T(X_1, \ldots, X_n)$

Definição 2.8 Consistência: *O estimador T_n diz-se (fracamente) consistente para $\tau(\theta)$ se, $\forall \theta \in \Theta$,*

$$\forall \varepsilon > 0 \lim_{n \to +\infty} P(|T_n - \tau(\theta)| > \varepsilon) = 0$$

ou seja, se $\forall \theta \in \Theta$, $T_n \xrightarrow{P} \tau(\theta)$. ■

Definição 2.9 Consistência em média quadrática: *O estimador T_n diz-se consistente em média quadrática para $\tau(\theta)$ se, $\forall \theta \in \Theta$,*

$$\lim_{n \to +\infty} E[(T_n - \tau(\theta))^2] = 0$$

ou seja, se $\forall \theta \in \Theta$, $T_n \xrightarrow{mq} \tau(\theta)$. ■

Observações:

- Pela desigualdade de Markov,

$$P(|T_n - \tau(\theta)| > \varepsilon) \leq E[(T_n - \tau(\theta))^2]/\varepsilon^2 = \mathrm{EMQ}_\theta(T_n)/\varepsilon^2$$

  logo consistência em média quadrática implica consistência (fraca)

- Como $E[(T_n - \tau(\theta))^2] = \mathrm{Var}(T_n) + [b_\theta(T)]^2$ segue-se que uma condição suficiente para que T_n seja consistente na estimação de $\tau(\theta)$ é que

$$\lim_{n\to+\infty} E[T_n] = \tau(\theta)$$
$$\lim_{n\to+\infty} \mathrm{Var}[T_n] = 0$$

- A consistência (mesmo em média quadrática) é uma propriedade pouco restritiva: se T_n é consistente, também o é $T_n^\star = \frac{n-a}{n-b} T_n$.

- Exemplo: a média amostral é um estimador consistente (em média quadrática, logo fracamente também) da média da população desde que esta exista.

2.3 Métodos de estimação

2.3.1 Método dos momentos

- Ideia: obter o estimador dos parâmetros da população igualando os momentos amostrais aos momentos populacionais
- Seja $\theta = (\theta_1, \ldots, \theta_k)$ o vector de parâmetros desconhecidos da população
- $\mu'_r = E[X^r]$ será uma função conhecida de θ: $\mu'_r = \psi_r(\theta_1, \ldots, \theta_k)$
- Considere-se os correspondentes momentos amostrais $M'_r = \sum_{i=1}^n X_i^r/n$ e forme-se o sistema de equações $M'_r = \psi_r(\theta_1, \ldots, \theta_k)$, $r = 1, \ldots, k$
- A solução deste sistema, que se assume existir e ser única, é designada por $\tilde{\Theta}_r = \varphi(X_1, \ldots, X_n)$ e dizemos que os estimadores $\tilde{\Theta}_1, \ldots, \tilde{\Theta}_k$ de θ foram obtidos pelo método dos momentos.

Exemplo 2.14 *Seja* $X_1, \ldots, X_n$ *uma amostra casual proveniente de uma população* $N(\mu, \sigma^2)$. *Sabemos que* $\mu = \mu_1'$ *e que* $\sigma^2 = \mu_2' - (\mu_1')^2$. *O sistema é então*

$$\begin{cases} M_1' = \mu_1' = \mu \\ M_2' = \mu_2' = \sigma^2 + \mu^2 \end{cases}$$

e a solução é que o estimador do método dos momentos de μ *é* $M_1' = \bar{X}$ *e que o de* σ^2 *é* $M_2' - (M_1')^2 = S^2$. ■

Exemplo 2.15 *Considere-se uma população* $X \sim G(\alpha, \lambda)$. *Sabemos que* $\mu_1' = \alpha/\lambda$ *e que* $Var(X) = \mu_2' - (\mu_1')^2 = \alpha/\lambda^2$. *Segue-se que o estimador do método dos momentos de* (α, λ) *é a solução do sistema*

$$\begin{cases} M_1' = \mu_1' = \alpha/\lambda \\ M_2' = \mu_2' = \alpha/\lambda^2 + (\alpha/\lambda)^2 \end{cases}$$

ou seja, $(\bar{X}^2/S^2, \bar{X}/S^2)$. ■

2.3.2 Método da máxima verosimilhança

- Ideia: propor como estimativa de θ o valor $\hat{\theta}$ que maximiza a função de verosimilhança.
- A estimativa de MV, se existir, é $\hat{\theta}$ tal que

$$L(\hat{\theta} \mid x_1, \ldots, x_n) \geq L(\theta \mid x_1, \ldots, x_n) \quad \forall \theta \in \Theta$$

- a correspondente variável aleatória determina o estimador de máxima verosimilhança.

Observações:

- na maior parte dos casos, é mais fácil maximizar o logaritmo da função de verosimilhança
- frequentemente, mas nem sempre, isso faz-se calculando os zeros da função score:

$$\frac{d \ln L(\theta \mid x_1, \ldots, x_n)}{d\theta} = 0$$

  e verificando que nesse ponto a segunda derivada é negativa
- Os estimadores de MV podem não ser únicos: amostra casual de dimensão n proveniente de população $\mathrm{U}(\theta - 1/2, \theta + 1/2)$, $\theta > 0$
- Nem sempre existe solução explícita, o que implica recorrer a métodos numéricos— e.g. função `nlm` no `R` (*non-linear minimization*)

Exemplo 2.16 $X_1, \ldots, X_n$ *amostra casual de dimensão* n *proveniente de uma população* $Po(\lambda)$

$$
\begin{aligned}
&L(\lambda \mid x_1, \ldots, x_n) \propto e^{-n\lambda}\lambda^{\sum x_i} \\
&\ln L(\lambda \mid x_1, \ldots, x_n) = c - n\lambda + n\bar{x}\ln\lambda \\
&\frac{d\ln L}{d\lambda} = -n + n\bar{x}/\lambda
\end{aligned}
$$

pelo que a função score tem um zero em $\hat{\lambda} = \bar{x}$. *É fácil verificar que se trata de um ponto de máximo de* $\ln L$, *pelo que o estimador de MV de* λ *é* $\bar{X}$. ■

Exemplo 2.17 $X_1, \ldots, X_n$ *amostra casual de dimensão* n *proveniente de uma população* $G(\alpha, \lambda)$ *com* λ *conhecido. Determinar estimativa de MV de* α:

$$
\begin{aligned}
&L(\alpha \mid x_1, \ldots, x_n) \propto \frac{\lambda^{n\alpha}}{\Gamma(\alpha)^n}\left(\prod x_i\right)^{\alpha-1} \\
&\ln L(\alpha \mid x_1, \ldots, x_n) = c + n\alpha\ln\lambda - n\ln\Gamma(\alpha) + (\alpha-1)\sum \ln x_i \\
&\frac{d\ln L}{d\alpha} = n\ln\lambda - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum \ln x_i
\end{aligned}
$$

os zeros da função score têm de ser calculados numericamente

Propriedades dos EMV:

- Se $\hat{\theta}$ for o estimador de MV de θ e se $\tau(\theta)$ for uma função biunívoca de θ, então $\tau(\hat{\theta})$ será estimador de MV de $\tau(\theta)$ — Invariância dos estimadores de MV que pode generalizar-se a situações onde $\tau(\theta)$ não é biunívoca
- Se T for suficiente, e se existir estimador de MV de θ, este é função de T
- O estimador mais eficiente para $\tau(\theta)$, se existir, é estimador de MV de $\tau(\theta)$
- Verificando-se determinadas condições de regularidade, o estimador de MV de θ, $\hat{\theta}$, é tal que

$$\sqrt{\mathcal{I}_{(X_1,\ldots,X_n)}(\theta)}(\hat{\theta}-\theta) \overset{a}{\sim} \mathrm{N}(0,1)$$

- No resultado acima é possível substituir $\mathcal{I}_{(X_1,\ldots,X_n)}(\theta)$
  - por $\mathcal{I}_{(X_1,\ldots,X_n)}(\hat{\theta})$, a informação de Fisher avaliada no estimador de MV
  - pela *informação de Fisher observada*

$$H(\theta) = -\frac{\partial^2 \ln L}{\partial \theta^2}(\hat{\theta} \mid x_1,\ldots,x_n)$$

  quantidades estas que não dependem de θ.

- O resultado acima garante que θ é consistente e assintoticamente mais eficiente