Métodos de estimação (pontual)

Prof. Caio Azevedo

Introdução

- Nosso objetivo é inferir (conjecturar a respeito do verdadeiro (ou verdadeiros) valor(es) de θ - possívelmente um vetor), ou seja, a respeito do parâmetro de interesse.
- Inferência Estatística: chegar à conclusões através de procedimentos estatísticos. Especificamente, com base em uma amostra, desejamos inferir a respeito do parâmetro θ associado à uma população de interesse.

Introdução

- **Inferência indutiva**: com base em sequências de premissas, chega-se a conclusões de interesse, através de nexos causais lógicas.
- **Inferência dedutiva**: com base em uma parte, obtem-se conclusões para o todo (esta é a que será abordada no curso).
- Primeiro passo: estimação pontual - obter uma estimativa, a mais acurada/precisa possível, com base em um estimador.
- Estimador: é uma estatística da qual se espera boas (ótimas) propriedades, em termos dos objetivos inferenciais.

Estrutura

- Seja $X_1, .., X_n$ uma aa de X, $X \sim f_X(.;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subset \mathcal{R}^k$. O objetivo é construir um estimador para $\boldsymbol{\tau}(\boldsymbol{\theta}) = (\tau_1(\boldsymbol{\theta}), ..., \tau_r(\boldsymbol{\theta}))'$, $r \leq k$
- Estimador: dizemos que $\widehat{\theta} \equiv \widehat{\theta}(\boldsymbol{X}) : \mathcal{X} \to B \subset \mathcal{R}$. Ou seja, um estimador é uma estatística (de preferência com o maior número de boas pripriedades possível), que tem por objetivo inferir a respeito de θ (espera-se que $B \subset \Theta$).
- Estimativa: $\widetilde{\theta} \equiv \widetilde{\theta}(\boldsymbol{x})$: valor numérico do estimador para uma amostra observada, $\boldsymbol{x} = (x_1, ..., x_n)'$.
- Podemos estar interessados em uma função de θ, digamos $\tau(\theta)$ ou num vetor (de funções) $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)'$ ou $(\boldsymbol{\tau}(\boldsymbol{\theta}) = (\tau_1(\boldsymbol{\theta}), ..., \tau_r(\theta))'$.

Características e qualidade dos estimadores

- Distribuição: $f_{\widehat{\theta}}(\widetilde{\theta};\theta)$.
- Valor esperado: $\mathcal{E}_{\theta}(\widehat{\theta})$ (mais próximo do verdadeiro valor, possível).
- Vício: $\mathcal{B}_{\theta}(\widehat{\theta}) = \mathcal{E}_{\theta}(\widehat{\theta} - \theta) = \mathcal{E}_{\theta}(\widehat{\theta}) - \theta$ (menor possível).
- Variância: $\mathcal{V}_{\theta}(\widehat{\theta})$ (menor possível).
- Erro quadrático médio (EQM):
  $\mathcal{EQM}_{\theta}(\widehat{\theta}) = \mathcal{E}[(\widehat{\theta} - \theta)^2] = \mathcal{B}_{\theta}(\widehat{\theta})^2 + \mathcal{V}_{\theta}(\widehat{\theta})$ (menor possível) (prova do resultado - exercício).
- Raiz quadrada do Erro quadrático médio:
  $\mathcal{REQM}_{\theta}(\widehat{\theta}) = \sqrt{\mathcal{EQM}_{\theta}(\widehat{\theta})}$ (menor possível).

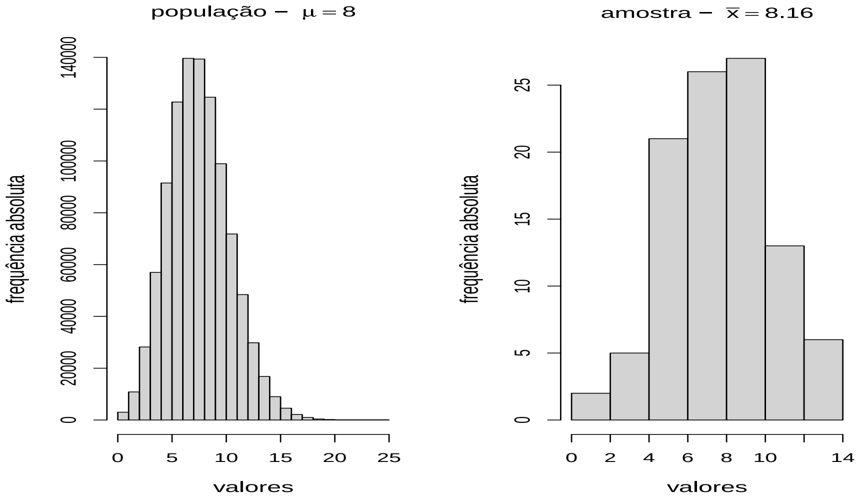
Características e qualidade dos estimadores

- Valor absoluto do vício relativo: $\mathcal{VAVR}_\theta(\widehat{\theta}) = \frac{|\mathcal{B}_\theta(\widehat{\theta})|}{|\theta|}$ (menor possível).
- Vício relativo: $\mathcal{VR}_\theta(\widehat{\theta}) = \frac{\mathcal{B}_\theta(\widehat{\theta})}{|\theta|}$ (menor possível).
- Consistência: $\widehat{\theta}_n \xrightarrow[n\to\infty]{P} \theta = \lim_{n\to\infty} P\left(|\widehat{\theta} - \theta| > \epsilon\right) = 0$
- Convergência em distribuição: $\widehat{\theta} \xrightarrow[n\to\infty]{D} D(\theta)$.

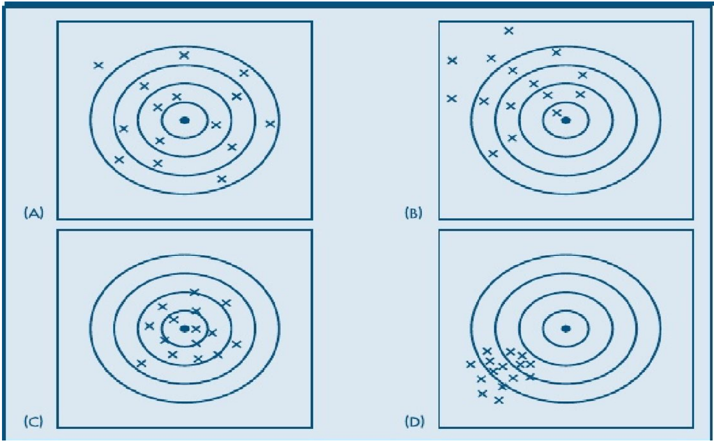
Características e qualidade dos estimadores

- Espera-se que estimadores sejam funções, caso existam, de estatísticas suficientes, completas e minimais, pois:
  - Suficiente: carrega toda a informação relevante.
  - Completa: não possui material ancilar (contaminação com informações que não dizem respeito ao parâmetro).
  - Minimal: máxima redução possível.
  - Dessa forma, é esperado que tais estimadores tenham boas (ótimas propriedades).
- Não faz sentido construir estimadores com base em estatísticas ancilares, pois elas não trazem informações a respeito do parâmetro.

População (N = 1.000.000, $\mu = 8$) e amostra (n = 100)



População e amostra: fonte



(A) - Muito acurado, mas pouco preciso; (B) Pouco acurado e pouco preciso; (C) Muito acurado e muito preciso; (D) Pouco acurado e muito preciso.

Métodos de estimação (ME)

- É impossível construir um estimador uniformemente ($\forall F_{\boldsymbol{X}}(.;\boldsymbol{\theta})$ e/ou $\forall \theta \in \Theta$) melhor do que todos os outros.
- Se $X_i \overset{iid}{\sim}$ Bernoulli(θ). Seja $\widehat{\theta} = 0,4$. Este estimador será o melhor sempre que $\theta = 0,4$ mas, provavelmente, será muito ruim $\forall\, \theta \neq 0,4$.
- ME: mecanismos que seguem alguma definição que (geralmente, com propriedades boas ou ótimas, em algum sentido).
- Dois tipos, basicamente:
  - Métodos genéricos baseados em critérios específicos: métodos dos momentos, máxima verossimilhança, mínimos quadrados.
  - Métodos que buscam estimadores ótimos: estimador não viciado de variância uniformemente mínima, estimador de Bayes, minimização do EQM.

Métodos dos momentos

- Baseia-se no fato, sob certas condições, de que

$$\frac{1}{n}\sum_{i=1}^{n} X_i^r \xrightarrow[n\to\infty]{P} \mathcal{E}(X^r), r = 1, 2, ...$$

- Deve-se, em relação à $\boldsymbol{\theta} \in \Theta \subset \Re^k$, resolver o seguinte sistema de equações:

$$\begin{aligned}
\mathcal{E}(X) &= \frac{1}{n}\sum_{i=1}^{n} X_i \\
\mathcal{E}(X^2) &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\
&\vdots \\
\mathcal{E}(X^k) &= \frac{1}{n}\sum_{i=1}^{n} X_i^k
\end{aligned}$$

Métodos dos momentos

- Principais características:
  - Não, necessariamente, leva à estimadores ótimos.
  - Em geral são facilmente obteníveis.
  - Nem sempre é possível obter distribuições/momentos, de forma exata.
  - Nem sempre os EMM são funções de estatísticas suficientes (completas e minimais).
  - Podem ser viciados.
  - As vezes precisamos utilizar momentos de ordem superior (k), pois momentos de ordem inferior ($< k$) podem não ser úteis.

Exemplo 1: *Bernoulli*(θ)

- Seja $X_1, ..., X_n$ uma aa de $X \sim$ Bernoulli(θ), $\theta \in (0,1)$.
- Temos que:

$$\mathcal{E}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i \to \widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

- Logo $\overline{X}$ é o EMM de θ.
- Além disso $\mathcal{E}(\widehat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{E}(X_i) = \theta$ (não viciado) e $\mathcal{V}(\widehat{\theta}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathcal{V}(X_i) = \frac{\theta}{n}$ $(\mathcal{V}(\widehat{\theta}) \xrightarrow[n\to\infty]{} 0)$.
- Além disso, $n\widehat{\theta} \sim$ binomial(n, θ).

Exemplo 2: $exp(\theta)$

- Seja $X_1, ..., X_n$ uma aa de $X \sim exp(\theta)$, $\theta \in (0, \infty)$ $(f_X(x; \theta) = \frac{1}{\theta} \exp\left\{-\frac{x}{\theta}\right\} \mathbb{1}_{(0,\infty)}(x))$.
- Temos que:

$$\mathcal{E}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i \to \widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

- Logo $\overline{X}$ é o EMM de θ.
- Além disso $\mathcal{E}(\widehat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{E}(X_i) = \theta$ (não viciado) e $\mathcal{V}(\widehat{\theta}) = \frac{1}{n^2}\sum_{i=1}^{n} \mathcal{V}(X_i) = \frac{\theta^2}{n}$ $(\mathcal{V}(\widehat{\theta}) \xrightarrow[n\to\infty]{} 0)$.
- Também, $\widehat{\theta} \sim gama(n, \theta/n)$ $(X \sim gama(a, b)$, $a, b \in (0, \infty)$, então $f_X(x; \theta) = \frac{1}{\Gamma(a)b} x^{a-1} \exp\left\{-\frac{x}{b}\right\} \mathbb{1}_{(0,\infty)}(x))$.

Exemplo 3: $N(\mu, \sigma^2)$

- Seja $X_1, ..., X_n$ uma aa de $X \sim N(\mu, \sigma^2)$, $\boldsymbol{\theta} = (\mu, \sigma^2)' \in \Re \times \Re^+$.
- Temos que:

$$\begin{cases} \mathcal{E}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i \to \widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} \\ \mathcal{E}(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \to \widehat{\mu}^2 + \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \\ \to \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \widehat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \end{cases}$$

- Assim, $\overline{X}$ e $\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ são, respectivamente, os EMM de μ e σ^2.
- Obs: usualmente usa-se a seguinte notação $\overline{X}_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$.

Exemplo 3: $N(\mu, \sigma^2)$

- Além disso, temos que $\overline{X} \sim N(\mu, \sigma^2/n)$ e $\frac{n\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-1)}$.
- Além disso:
  - $\mathcal{E}(\widehat{\mu}) = \mu$ (não viciado) e $\mathcal{V}(\widehat{\mu}) = \frac{\sigma^2}{n}$ ($\mathcal{V}(\widehat{\mu}) \xrightarrow[n\to\infty]{} 0$),
  - $\mathcal{E}(\widehat{\sigma}^2) = \frac{n-1}{n}\sigma^2$ (viciado), $\mathcal{B}(\widehat{\sigma}^2) = \frac{-\sigma^2}{n}$ ($\mathcal{B}(\widehat{\sigma}^2) \xrightarrow[n\to\infty]{} 0$), $\mathcal{V}(\widehat{\sigma}^2) = \frac{2(n-1)}{n^2}(\sigma^2)^2$ ($\mathcal{V}(\widehat{\sigma}^2) \xrightarrow[n\to\infty]{} 0$) e $\mathcal{EQM}(\sigma^2) = \frac{(2n-1)(\sigma^2)^2}{n^2}(\mathcal{EQM}(\widehat{\sigma}^2) \xrightarrow[n\to\infty]{} 0)$.

Exemplo 4: $U(0, \theta)$

- Seja $X_1, ..., X_n$ uma aa de $X \sim U(0, \theta)$, $\boldsymbol{\theta} \in \Re^+$
- Temos que:

$$\mathcal{E}(X) = \frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow \frac{\widehat{\theta}}{2} = \overline{X} \rightarrow \widehat{\theta} = 2\overline{X}$$

- Além disso, temos que $\mathcal{E}(\widehat{\theta}) = 2\frac{\theta}{2} = \theta$ (não viciado) e $\mathcal{V}(\widehat{\theta}) = \frac{4\theta^2}{12n^2} = \frac{\theta^2}{3n^2}$ $(\mathcal{V}(\widehat{\theta}) \xrightarrow[n \to \infty]{} 0)$.

Exemplo 4: $U(-\theta, \theta)$

- Seja $X_1, ..., X_n$ uma aa de $X \sim U(-\theta, \theta)$, $\boldsymbol{\theta} \in \Re^+$
- Neste caso, note que $\mathcal{E}(X) = 0$, portanto, esta não é útil para obter o EMM de θ. Por outro lado, $\mathcal{E}(X^2) = \frac{\theta^2}{3}$. Assim, temos que:

$$\mathcal{E}(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \rightarrow \frac{\widehat{\theta}^2}{3} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \rightarrow \widehat{\theta} = \sqrt{\frac{3}{n}\sum_{i=1}^{n} X_i^2} = \sqrt{3\overline{X}_2}$$

- Nesse caso, os cálculos de $\mathcal{E}(\widehat{\theta})$ e $\mathcal{V}(\widehat{\theta})$ são complicados.
- Exercício: obter os EMM para θ no caso em que $X \sim Poisson(\theta)$ e para $\boldsymbol{\theta} = (a, b)'$, no caso em que $X \sim gama(a, b)$. Calcule, se viável, as distribuições exatas, bem como $\mathcal{E}(.)$, $\mathcal{V}(.)$, $\mathcal{B}(.)$ e $\mathcal{EQM}(.)$.

Métodos dos momentos

- Resultado assintótico: Seja $h(\boldsymbol{\theta}) = (h_1(\boldsymbol{\theta}), ..., h_k(\boldsymbol{\theta}))'$, $h_r(\boldsymbol{\theta}) = \mathcal{E}(X^r)$, $r = 1, 2, .., k$ e suponhamos que

$$\boldsymbol{H}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial h_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial h_1(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial h_1(\boldsymbol{\theta})}{\partial \theta_k} \\ \frac{\partial h_2(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial h_2(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial h_2(\boldsymbol{\theta})}{\partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_k(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial h_k(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial h_k(\boldsymbol{\theta})}{\partial \theta_k} \end{bmatrix}$$

seja de posto completo e que $\frac{\partial}{\partial \theta_i} h_j(\boldsymbol{\theta})$, $i, j = 1, 2, ..., k$ seja contínua em θ.

Métodos dos momentos

- Então,

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow[n\to\infty]{D} N_k\left(\mathbf{0}, \left[\boldsymbol{H}^{-1}(\boldsymbol{\theta})\right]' \boldsymbol{\Sigma} \left[\boldsymbol{H}^{-1}(\boldsymbol{\theta})\right]\right).$$

- Em que $\boldsymbol{\Sigma}_{i,j} = [\mu_{i+j} - \mu_i \mu_j]$, em que $\mu_i = \mathcal{E}(X^i), \forall i$. Demonstração: aqui.

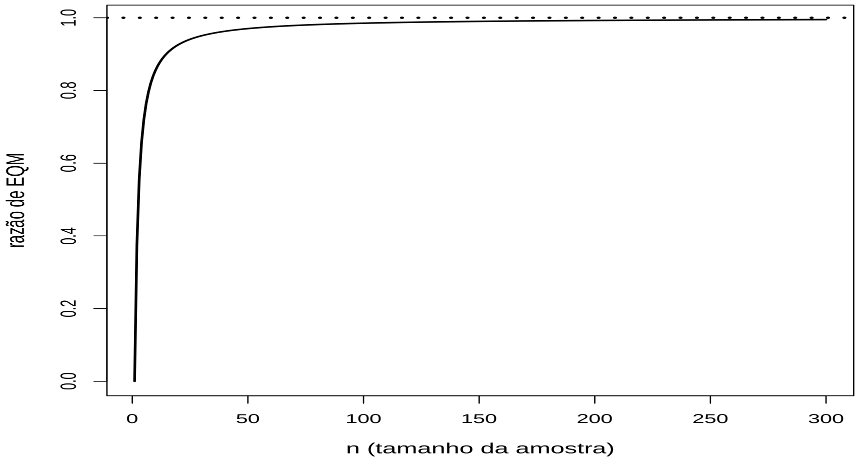
Exemplo de comparação de estimadores

- Seja $X_1, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, $\boldsymbol{\theta} = (\mu, \sigma^2)'$. Considere o interesse em estimar σ^2, através de um dos dois seguintes estimadores: $\widehat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$ e $\widehat{\sigma}_2^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2$.
- Note que $\widehat{\sigma}_1^2 = \frac{n-1}{n}\widehat{\sigma}_2^2$. Assim, $\lim_{n\to} \frac{\widehat{\sigma}_1^2}{\widehat{\sigma}_2^2} = \lim_{n\to} \frac{n-1}{n} = 1$(assintoticamente equivalentes).
- Sabemos que $\frac{n-1}{\sigma^2}\widehat{\sigma}_2^2 \sim \chi^2_{(n-1)}$.

Exemplo de comparação de estimadores

- Logo, $\mathcal{E}(\widehat{\sigma}_1^2) = \frac{n-1}{n}\sigma^2$ e $\mathcal{E}(\widehat{\sigma}_2^2) = \sigma^2$.
- Por outro lado, $\mathcal{B}(\widehat{\sigma}_1^2) = -\frac{1}{n}\sigma^2$ e $\mathcal{B}(\widehat{\sigma}_2^2) = 0$.
- Além disso, $\mathcal{V}(\widehat{\sigma}_1^2) = (n-1)\frac{2(\sigma^2)^2}{n^2}$ e $\mathcal{V}(\widehat{\sigma}_2^2) = \frac{2(\sigma^2)^2}{n-1}$.
- Também, $\mathcal{EQM}(\widehat{\sigma}_1^2) = \frac{2n-1}{n^2}(\sigma^2)^2$ e $\mathcal{EQM}(\widehat{\sigma}_2^2) = \mathcal{V}(\widehat{\sigma}_2^2) = \frac{2(\sigma^2)^2}{n-1}$.
- Portanto, note que $\frac{\mathcal{EQM}(\widehat{\sigma}_1^2)}{\mathcal{EQM}(\widehat{\sigma}_2^2)} = \frac{(2n-1)(n-1)}{2n^2} < 1 \rightarrow n > 1/3$.

Comparação dos EQM's



Método de máxima verossimilhança

- Vimos anteriormente que, para uma aa de tamanho n de X, de sorte que $f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) = \prod_{i=1}^{n} f_X(x_i;\boldsymbol{\theta})$, a verossimilhança associada a $\boldsymbol{\theta} \in \Theta$ é dada por

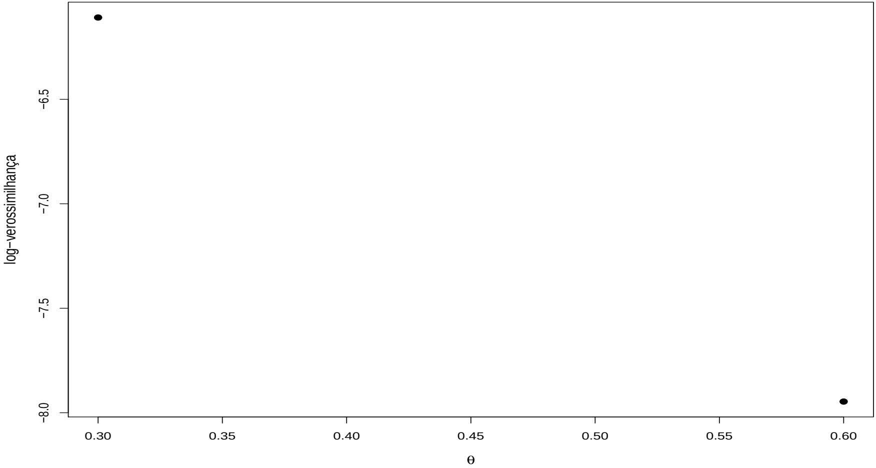
$$L(\boldsymbol{\theta};\boldsymbol{x}) \equiv L(\boldsymbol{\theta}) = f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) = \prod_{i=1}^{n} f_X(x_i;\boldsymbol{\theta}).$$

- A verossimilhança nos fornece, então, a probabilidade de observar uma dada amostra em função do θ de forma exata (caso discreto) ou de forma aproximada (caso contínuo). Em particular, a amostra que fora observada.
- Assim, se obtivermos, através de algum procedimento, o máximo da verossimilhança, em termos de θ, obteremos o valor mais provável responsável pela geração de uma dada amostra.

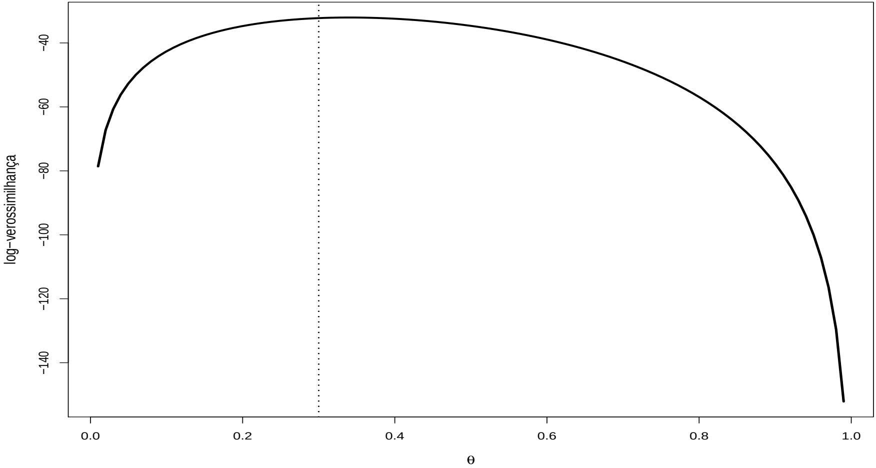
Métodos de máxima verossimilhança

- Exemplo (Bernoulli): Seja $X_1, ..., X_n \stackrel{iid}{\sim}$ Bernoulli(θ). Considere duas situações:
  1. $n = 10$ $\boldsymbol{x} = (0, 0, 1, 0, 0, 1, 0, 0, 1, 0)'$, $\Theta = \{0,3; 0,6\}$.
  2. n=100 valores simulados de uma Bernoulli$(0,3)$, $\Theta = (0,1)$.
- Exemplo (exponencial): Seja $X_1, ..., X_n \stackrel{iid}{\sim} \exp(\theta), \mathcal{E}(X) = \theta$. Considere uma amostra com n=100 valores simulados de uma $\exp(10)$.
- Nos três slides a seguir, mostraremos as respectivas verossimilhanças.

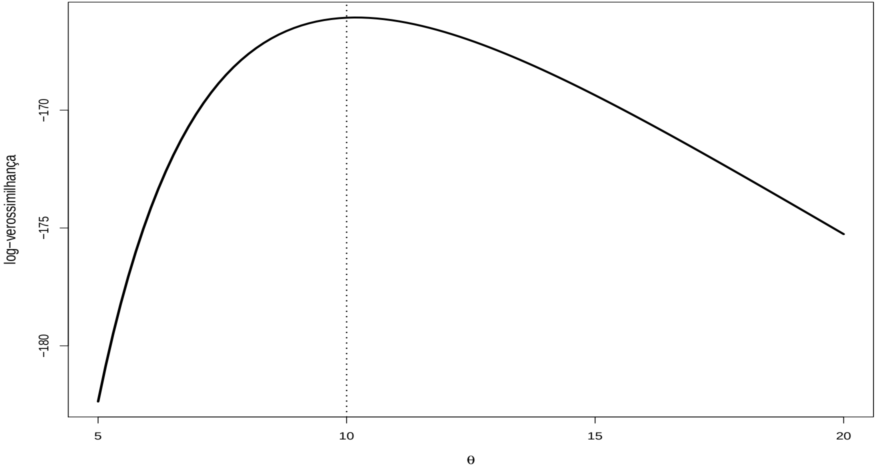
Exemplo 1 (Bernoulli)



Exemplo 2 (Bernoulli)



Exemplo exponencial



Método de máxima verossimilhança

- Significado(s) da verosimilhança:
  - Caso discreto:

$$f_{\boldsymbol{X}}(\boldsymbol{x};\theta) = P_\theta(\boldsymbol{X} = \boldsymbol{x}) = P_\theta(X_1 = x_1, X_2 = x_2, ..., X_n = x_n).$$

  - Caso contínuo:

$$f_{\boldsymbol{X}}(\boldsymbol{x};\theta) =$$
$$\lim_{\Delta \to 0} P_\theta(x_1 - \Delta \leq X_1 \leq x_1 + \Delta, ..., x_n - \Delta \leq X_n \leq x_n + \Delta).$$

Método de máxima verossimilhança

- Def: Estimador de máxima verossimilhança (EMV) - Seja $X_1, ..., X_n$ va's (não necessariamente iid), $f_{\boldsymbol{X}}(.;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subseteq \mathcal{R}^k$, é a fdp conjunta de $\boldsymbol{X}$. Qualquer valor $\widetilde{\boldsymbol{\theta}} = \widetilde{\boldsymbol{\theta}}(\boldsymbol{x})$, tal que:

$$L(\widetilde{\boldsymbol{\theta}}) = \sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}),$$

  é chamado de estimador de máxima verossimilhança de $\boldsymbol{\theta}$ (pode haver somente um, mais de um, infinitos ou nenhum).
- Assim, a respectiva quantidade $\widehat{\boldsymbol{\theta}} \equiv \widehat{\boldsymbol{\theta}}(\boldsymbol{X})$ é o (um) estimador de máxima versossimilhança de $\boldsymbol{\theta}$.

Método de máxima verossimilhança

- Note que, $\forall \widetilde{\boldsymbol{\theta}}$ emv, temos que :
  - $L(\widetilde{\boldsymbol{\theta}}; \boldsymbol{x}) > L(\boldsymbol{\theta}; \boldsymbol{x})$, $\forall \boldsymbol{\theta} \in \Theta, \boldsymbol{\theta} \neq \widetilde{\boldsymbol{\theta}}$.
  - $L(\widetilde{\boldsymbol{\theta}}; \boldsymbol{x}) \geq L(\boldsymbol{\theta}; \boldsymbol{x}), \forall \boldsymbol{\theta} \in \Theta$.
- Em geral $\sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}) = \max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta})$.
- Muitas vezes, dispositivos usuais de maximização (solução de sistemas de equações lineares ou otimização numérica) podem/devem ser utilizados.

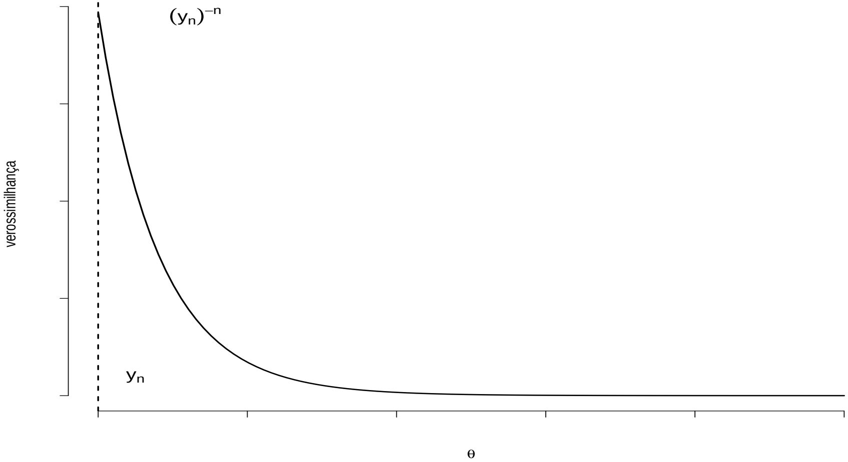
Exemplo

- Exemplo: Seja $X_1, ..., X_n$ uma aa de $X \sim U(0, \theta), \theta \in \mathcal{R}^+$. Então:

$$f_{\boldsymbol{X}}(\boldsymbol{x}; \theta) = \frac{1}{\theta^n} \mathbb{1}_{(0,\theta)}(y_n) \mathbb{1}_{(0,y_n)}(y_1) \propto \frac{1}{\theta^n} \mathbb{1}_{(y_n,\infty)}(\theta)$$

- Como $L(\theta)$ não é contínua em θ, não é possível derivá-la para se obter seu máximo. Contudo, note que, graficamente (próximo slide):
- Podemos notar, então, que $\frac{1}{y_n^n} > \frac{1}{\theta^n}$, $\forall \theta \in [y_n, \infty)$. Portanto, $\widehat{\theta} = Y_n$ é o estimador de máxima verossimilhança de θ.

Exemplo $U(0, \theta)$



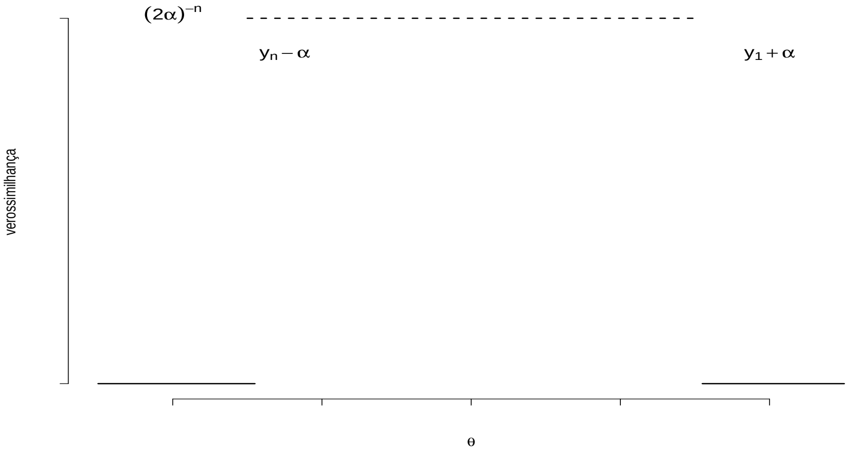
Métodos de máxima verossimilhança

- Exemplo 2: Seja $X_1, ..., X_n$ yma aa de $X \sim U(\theta - \alpha, \theta + \alpha)$, $\alpha > 0$ (conhecido), $\theta \in \mathcal{R}$. Calcular o emv de θ.
- Temos que:

$$L(\theta) = \frac{1}{(2\alpha)^n} \mathbb{1}_{(y_n - \alpha, y_1 + \alpha)}(\theta)$$

- No gráfico do slide seguinte, vemos que $L(\widetilde{\theta}) = \sup_{\theta \in \Theta} L(\theta), \forall \widetilde{\theta} \in (y_n - \alpha, y_1 + \alpha)$, $\exists$ inifinitas estimativas de máxima verossimilhança de θ.
- Assim $\widehat{\theta}$, $\forall$ $Y_n - \alpha < \widehat{\theta} < Y_1 - \alpha$, será um EMV de θ, particularmente $\widehat{\theta} = \frac{Y_1 + Y_n}{2}$.

Exemplo $U(\theta - \alpha, \theta + \alpha)$



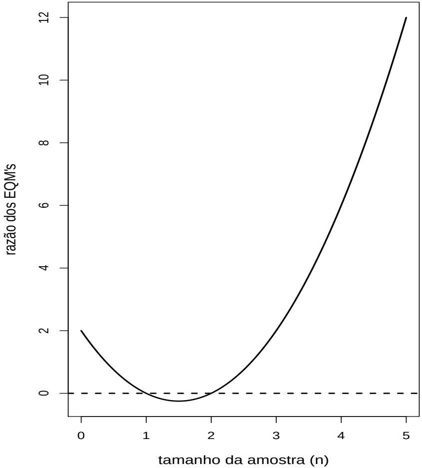
Comparação entre $\widehat{\theta}_{MM}$ e $\widehat{\theta}_{MV}$ no caso de $U(0,\theta)$

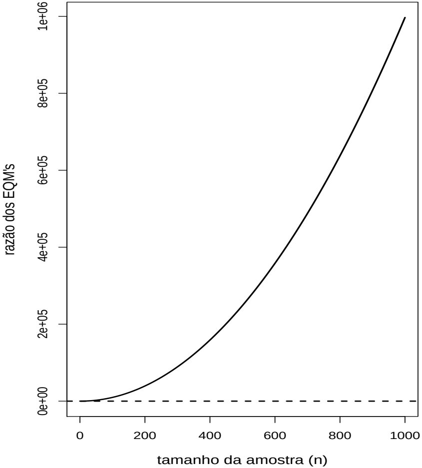
- Lembremos que $\mathcal{E}(X) = \frac{\theta}{2}$ e $\mathcal{V}(X) = \frac{\theta^2}{12}$.
- $\widehat{\theta}_{MM} = 2\overline{X}$ e $\widehat{\theta}_{MV} = Y_n = \max(\boldsymbol{X})$.
- Distribuições: $\widehat{\theta}_{MM}$ (obter via fgm), $f_{Y_n}(y;\theta) = \frac{ny^{n-1}}{\theta^n}\mathbb{1}_{(0,\theta)}(y)$.
- Esperança: $\mathcal{E}(\widehat{\theta}_{MM}) = \theta$ e $\mathcal{E}(\widehat{\theta}_{MV}) = \frac{n}{n+1}\theta$.
- Vício: $\mathcal{B}(\widehat{\theta}_{MM}) = 0$ e $\mathcal{B}(\widehat{\theta}_{MV}) = -\frac{\theta}{n+1}$.
- Variância: $\mathcal{V}(\widehat{\theta}_{MM}) = \frac{\theta^2}{3n}$ e $\mathcal{V}(\widehat{\theta}_{MV}) = \frac{n\theta^2}{(n+1)^2(n+2)}$.

Comparação entre $\widehat{\theta}_{MM}$ e $\widehat{\theta}_{MV}$ no caso de $U(0,\theta)$

- EQM: $\mathcal{EQM}(\widehat{\theta}_{MM}) = \mathcal{V}(\widehat{\theta}_{MM}) = \frac{\theta^2}{3n}$ e $\mathcal{EQM}(\widehat{\theta}_{MV}) = \mathcal{V}(\widehat{\theta}_{MV}) + \mathcal{B}^2(\widehat{\theta}_{MV}) = \frac{2\theta^2}{(n+2)(n+1)}$.
- Assim $\frac{\mathcal{EQM}(\widehat{\theta}_{MM})}{\mathcal{EQM}(\widehat{\theta}_{MV})} = \frac{(n+1)(n+2)}{6n} > 1 \leftrightarrow n^2 - 3n + 2 > 0$ (raízes $n_1 = 1$ e $n_2 = 2$). Neste caso, a desigualdade será satisfeita sempre que $n < 1$ ou $n > 2$. Ou seja, $\widehat{\theta}_{MV}$ é preferível a $\widehat{\theta}_{MM}$, segundo o EQM.
- Note que podemos obter um estimador não viciado de θ a partir de $\widehat{\theta}_{MV}$, fazendo $\widehat{\theta} = \frac{n+1}{n}\widehat{\theta}_{MV}$.
- Assim, $\mathcal{V}(\widehat{\theta}) = \frac{(n+1)^2}{n^2}\frac{n\theta^2}{(n+1)^2(n+2)} = \frac{\theta^2}{n(n+2)}$.
- Portanto $\frac{\mathcal{V}(\widehat{\theta})}{\mathcal{V}(\widehat{\theta}_{MM})} = \frac{3}{n+2} < 1 \rightarrow n > 1$. Assim, segundo a variância, $\widehat{\theta}$ também é preferível a $\widehat{\theta}_{MM}$.

Comparação de estimadores $U(0, \theta)$





Resultado

- Seja $\Theta \subseteq \mathcal{R}$ e $L(\theta)$ duas vezes diferenciável, tal que

$$\left.\frac{\partial L(\theta)}{\partial \theta}\right|_{\theta=\widetilde{\theta}} = 0; \left.\frac{\partial^2 L(\theta)}{\partial \theta^2}\right|_{\theta=\widetilde{\theta}} < 0$$

  então $\widetilde{\theta}$ é a emv de θ e, assim $\widehat{\theta}$ é o EMV de θ.
- OBS: $\widetilde{\theta}$ maximiza $L(\theta) \leftrightarrow \widetilde{\theta}$ maximiza $l(\theta) = \ln L(\theta) \leftrightarrow$, pois $\ln(.)$ é uma função monótona e $L(\theta) > 0 \ \forall \theta \in \Theta$ e $\ln(L(\theta))$ é chamada de logverossimilhança.

Exemplo

- Exemplo 3: Seja $X_1, ..., X_n$ uma aa de X, $X \sim$ Bernoulli(θ), $\theta \in [0, 1]$. Temos que:

$$
\begin{aligned}
L(\theta) &= \theta^{n\overline{x}}(1-\theta)^{n(1-\overline{x})}\mathbb{1}_{\{0,1\}^n}(\boldsymbol{x}) \rightarrow \\
&\rightarrow l(\theta) = n\overline{x}\ln(\theta) + n(1-\overline{x})\ln(1-\theta) + \text{const.}
\end{aligned}
$$

- Logo, a função escore ($S(\theta)$) é dada por:

$$S(\theta) = \frac{dl(\theta)}{d\theta} = \frac{n\overline{x}}{\theta} - \frac{n(1-\overline{x})}{1-\theta}$$

Assim, $S(\widetilde{\theta}) = 0 \rightarrow (1-\widetilde{\theta})\overline{x} = \widetilde{\theta}(1-\overline{x}) \rightarrow \widetilde{\theta} = \overline{x}$

Exemplo

- Por outro lado, a função Hessiana ($H(\theta)$) é dada por

$$
\begin{aligned}
H(\theta) &= \frac{d^2 l(\theta)}{d\theta^2} = -\frac{n\overline{x}}{\theta^2} - \frac{n(1-\overline{x})}{(1-\theta)^2} \\
&= -n\left[\frac{\overline{x}}{\theta^2} + \frac{1-\overline{x}}{(1-\theta)^2}\right]
\end{aligned}
$$

Logo, $H(\widetilde{\theta}) = -n\left[\frac{1}{\widetilde{\theta}} + \frac{1}{1-\widetilde{\theta}}\right] < 0$. Assim, $\widetilde{\theta}$ é a emv de θ e, portanto, $\widehat{\theta}$ é o emv de θ.

Exemplo

- Note que $\mathcal{E}(\widehat{\theta}) = \theta$ e $\mathcal{V}(\widehat{\theta}) = \frac{\theta(1-\theta)}{n}$ (veja os slides) para ver algumas propriedades desse estimador.
- Contudo, a distribuição exata de $\widehat{\theta}$ é complicada de ser obtida.
- Exercício. Prove que, se $a) \, x_i = 0, \forall i$ e $b) \, x_i = 1, \forall i$ então, respectivamente, $\widehat{\theta}_{MV} = 0$ e $\widehat{\theta}_{MV} = 1$.

Resultado

- Suponha agora que $\Theta \subset \mathcal{R}^k$ e que $l(\boldsymbol{\theta})$ seja (duplamente: derivadas de primeira e segunda ordem) diferenciável com respeito à $\boldsymbol{\theta}$. Seja $\widetilde{\boldsymbol{\theta}}$ **o valor** de $\boldsymbol{\theta}$ que satisfaz:

$$\begin{cases} \dfrac{\partial l(\boldsymbol{\theta})}{\partial \theta_1} = 0 \\ \dfrac{\partial l(\boldsymbol{\theta})}{\partial \theta_2} = 0 \\ \vdots \\ \dfrac{\partial l(\boldsymbol{\theta})}{\partial \theta_k} = 0 \end{cases} \tag{2}$$

então $\widetilde{\boldsymbol{\theta}}$ é um candidato à emv de θ.

Resultado

- Se além da condição (5) ser satisfeita, tivermos (k=2) que:

$$\left(\left. \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \theta_1^2} \right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \right) \left(\left. \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \theta_2^2} \right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \right) - \left(\left. \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2} \right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} \right)^2 > 0$$

e

$$\left. \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \theta_1^2} \right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} < 0 \text{ ou } \left. \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \theta_2^2} \right|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}} < 0$$

então $\widetilde{\boldsymbol{\theta}}$ é a emv de $\boldsymbol{\theta}$ e, consequentemente, $\widehat{\boldsymbol{\theta}}$ é o emv de $\boldsymbol{\theta}$.

Resultado

- No caso geral ($k \geq 2$), devemos ter que $\boldsymbol{H}(\boldsymbol{\theta})$

$$\boldsymbol{H}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_1^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_1 \theta_2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_1 \theta_p} \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_2 \theta_1} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_2 \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_p \theta_1} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_p \theta_2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_p^2} \end{bmatrix}$$

  seja negativa definida (veja aqui).

- Neste caso ($k \geq 2$) não, necessariamente, o ponto encontrado é de máximo.

Exemplo

- Exemplo 4: Seja $X_1, ..., X_n$ uma aa de $X \sim N(\mu, \sigma^2)$, $\boldsymbol{\theta} = (\mu, \sigma^2)$ (desconhecido), $\Theta = \mathcal{R} \times \mathcal{R}^2$.
- Temos que

$$L(\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ -\frac{-1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\} \mathbb{1}_{\mathcal{R}^n}(\boldsymbol{x})$$

$$\rightarrow \quad l(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 - \frac{n}{2} \ln(\sigma^2) + \text{const}$$

- Assim, vem que o vertor escore é dado por $\boldsymbol{S} = (S(\mu), S(\sigma^2))'$, em que:

$$S(\mu) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) \, ; \; S(\sigma^2) = \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (x_i - \mu)^2 - \frac{n}{2\sigma^2}$$

Exemplo

- Portanto, de $\boldsymbol{S}(\widetilde{\boldsymbol{\theta}}) = \boldsymbol{0}$, vem que

$$\begin{cases} \frac{1}{\widetilde{\sigma}^2}\sum_{i=1}^{n}(x_i - \widetilde{\mu}) = 0 \ (*) \\ \frac{1}{2(\widetilde{\sigma}^2)^2}\sum_{i=1}^{n}(x_i - \widetilde{\mu})^2 = \frac{n}{2\widetilde{\sigma}^2} \ (**) \end{cases}$$

De (*) temos que $\widetilde{\mu} = \overline{x} \, (***)$ e, assim, de $(***)$ em (**), vem que $\widetilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \widetilde{\mu})^2$.

- Por outro lado, temos que

$$\boldsymbol{H}(\boldsymbol{\theta}) = \begin{bmatrix} H(\mu,\mu) & H(\mu,\sigma^2) \\ H(\sigma^2,\mu) & H(\sigma^2,\sigma^2) \end{bmatrix}$$

Exemplo

- Cujos elementos são dados por:

$$H(\mu,\mu) = -\frac{n}{\sigma^2} \; ; H(\sigma^2,\sigma^2) = \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$H(\mu,\sigma^2) = H(\sigma^2,\mu) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

Exemplo

- Logo

$$H(\widetilde{\mu},\widetilde{\mu}) = -\frac{n}{\widetilde{\sigma}^2}$$

$$H(\widetilde{\sigma}^2,\widetilde{\sigma}^2) = \frac{n}{2(\widetilde{\sigma}^2)^2} - \frac{1}{(\widetilde{\sigma}^2)^3}\sum_{i=1}^{n}(x_i-\widetilde{\mu})^2 = \frac{n}{2(\widetilde{\sigma}^2)^2} - \frac{n}{(\widetilde{\sigma}^2)^2} = -\frac{n}{2(\widetilde{\sigma}^2)^2}$$

$$H(\widetilde{\mu},\widetilde{\sigma}^2) = H(\widetilde{\sigma}^2,\widetilde{\mu}) = -\frac{1}{2\widetilde{\sigma}^2}\sum_{i=1}^{n}(x_i-\widetilde{\mu}) = 0$$

Exemplo

- Por outro lado, temos que $H(\widetilde{\mu},\widetilde{\mu})H(\widetilde{\sigma}^2,\widetilde{\sigma}^2) - [H(\widetilde{\mu},\widetilde{\sigma}^2)]^2 = \frac{n^2}{2(\widetilde{\sigma}^2)^3} > 0$ e $H(\widetilde{\mu},\widetilde{\mu}) = -\frac{n}{\widetilde{\sigma}^2} < 0$
- Portanto, conclui-se que $\widetilde{\mu}$ e $\widetilde{\sigma}^2$ são as emv de $\boldsymbol{\theta}$ e, assim $\widehat{\mu} = \overline{X}$ e $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ são os respectivos EMV.
- Exercício: obtenha os emv de $\boldsymbol{\theta}$, baseado em uma aa de X, quando :
  a) $X \sim U(\alpha,\beta)$, $\boldsymbol{\theta} = (\alpha,\beta)'$
  b) $X \sim U(\alpha-\gamma,\alpha+\gamma)$, $\boldsymbol{\theta} = (\alpha,\gamma)'$
  considerando i) α conhecido, ii) γ conhecido e iii) ambos desconhecidos

Exemplo

- Exemplo 5: Seja X uma va com distribuição HG (N, n, a) (HG $\equiv$ hipergeométrica) em que N: tamanho da população (muitas vezes o parâmetro de interesse), n: tamanho da amostra, a: número de elementos com a caraterística desejada, na população (também pode ser um parâmetro de interesse).
- Temos que (N será o parâmetro de interesse):

$$f_X(x;N) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}} \mathbb{1}_{\{c_1,\ldots,c_2\}}(x)$$

em que $c_1 = \max(0, n - N + a)$ e $c_2 = \min(n, a)$.

Exemplo

- Defina $r(N) = \frac{L(N)}{L(N-1)}$, logo

$$r(N) = \frac{\frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}}{\frac{\binom{a}{x}\binom{N-1-a}{n-x}}{\binom{N-1}{n}}} = \frac{(N-a)(N-n)}{N(N-a-n+x)}$$

- Note, então, que

$$\begin{aligned} r(N) &> 1 \leftrightarrow (N-a)(N-n) > N(N-a-n+x) \leftrightarrow \\ & N^2 - Nn - Na - an > N^2 - Na - Nn + Nx \rightarrow N > \frac{an}{x} \end{aligned}$$

Analogamente, se $r(N) < 1 \rightarrow N < \frac{an}{x}$.

Exemplo

- Então, a emv de N será, aproximadamente $\frac{an}{x}$, ou seja $\widetilde{N} \approx \frac{an}{x}$.
- Se $\frac{an}{x}$ não for inteiro, podemos considerar $\lceil \frac{an}{x} \rceil$ e $\lfloor \frac{an}{x} \rfloor$ e avaliar onde a verossimilhança é maior.
- Consequentemente, o emv de N será $\widehat{N} = \frac{an}{X}$.

Exemplo

- Exemplo 6: Seja $X_1, ..., X_n$ uma aa de X, $X \sim \text{gama}(r, \lambda)$, $\mathcal{E}(X) = r\lambda$, $\boldsymbol{\theta} = (r, \lambda)'$, $\Theta = \mathcal{R}^+ \times \mathcal{R}^+$.
- Temos que

$$\begin{aligned} L(\boldsymbol{\theta}) &= \frac{1}{\lambda^{nr}(\Gamma(r))^n} e^{-\frac{n}{\lambda}\overline{x}} \prod_{i=1}^{n} x_i^{r-1} \mathbb{1}_{(\mathcal{R}^+)^n}(\boldsymbol{x}) \\ \rightarrow & \quad l(\boldsymbol{\theta}) = -\frac{1}{\lambda} n\overline{x} + (r-1)\sum_{i=1}^{n} \ln(x_i) - nr\ln(\lambda) - n\ln(\Gamma(r)) \end{aligned}$$

Exemplo

- Portanto, as componentes do vetor escore, são dadas por (função digama $\Psi(r) = \frac{\Gamma'(r)}{\Gamma(r)}$)

$$S(r) = \sum_{i=1}^{n} \ln(x_i) - n\ln(\lambda) - n\Psi(r) \; ; S(\lambda) = \frac{n\overline{x}}{\lambda^2} - \frac{nr}{\lambda}$$

- Dessa forma,

$$\begin{cases} \sum_{i=1}^{n} \ln(x_i) - n\ln(\widetilde{\lambda}) - n\Psi(\widetilde{r}) = 0 \\ \frac{n\overline{x}}{\widetilde{\lambda}^2} - \frac{n\widetilde{r}}{\widetilde{\lambda}} \quad = 0 \end{cases}$$

$$\rightarrow \quad \begin{cases} n\ln(\widetilde{\lambda}) + n\Psi(\widetilde{r}) = \sum_{i=1}^{n} \ln(x_i) \; (*) \\ \widetilde{\lambda} = \frac{\overline{x}}{\widetilde{r}} (**) \end{cases}$$

Exemplo

- De (**) em (*), temos que:

$$-n\ln(\widetilde{r}) + n\Psi(\widetilde{r}) = \sum_{i=1}^{n}\ln(x_i) - n\ln(\overline{x})$$

a qual não tem solução explícita e, assim, algum método numérico para resolução de equações (sistemas) não linear(es), tem de ser empregado, como: algoritmo de Newton-Raphson, Escore de Fisher, BFGS, Nelder-Mead, L-BFGS-B, gradiente conjugado, simulated annealing entre outros. Veja aqui e aqui, por exemplo.

Exemplo

- Neste caso, a determinação se as estimativa são de mv, deve ser feita por métodos numéricos (avaliando se a matriz Hessiana é negativa definida)
- Caso r seja conhecido, pode-se provar que $\widehat{\lambda} = \frac{\overline{X}}{r}$ é o emv de λ (exercício). Obtenha também $\mathcal{E}(\widehat{\lambda})$, $\mathcal{V}(\widehat{\lambda})$ bem como sua distribuição exata (compare como respectivo emm).

Propriedade da invariância dos emv

- Seja $\Theta \subseteq \mathcal{R}^k$ e $\boldsymbol{\tau} : \Theta \to \Lambda \subseteq \mathcal{R}^r$, em que $\boldsymbol{\tau}(\boldsymbol{\theta}) = (\tau_1(\boldsymbol{\theta}), ..., \tau_r(\boldsymbol{\theta}))'$, $r \leq k$.
- Suponha que $r = k = 1$ e $\tau(\theta) = \eta$ uma transformação 1 a 1, temos que

$$L^*(\eta, \boldsymbol{x}) = L(\tau^{-1}(\eta), \boldsymbol{x}) = L(\theta, \boldsymbol{x}) = \prod_{i=1}^{n} f_X(x_i; \theta)$$

- Assim, podemos verificar que:

$$\sup_{\eta \in \Lambda} L^*(\eta, \boldsymbol{x}) = \sup_{\eta \in \Lambda} L(\tau^{-1}(\eta); \boldsymbol{x}) = \sup_{\theta \in \Theta} L(\theta; \boldsymbol{x})$$

- Portanto, $\sup_{\eta \in \Lambda} L^*(\eta, \boldsymbol{x}) = L(\widetilde{\eta}; \boldsymbol{x})$, $\widetilde{\eta} = \tau(\widetilde{\theta})$.
- Logo, $\widehat{\eta} = \tau(\widehat{\theta})$ é o emv de η.

Teorema

- Princípio da invariância dos emv: seja $\widehat{\boldsymbol{\theta}}$ o emv de $\boldsymbol{\theta} \in \Theta \subseteq \mathcal{R}^k$ e $\boldsymbol{\tau} : \Theta \to \Lambda \subseteq \mathcal{R}^r, r \leq k$ (não, necessariamente, uma função 1 a 1). Então $\tau(\widehat{\boldsymbol{\theta}})$ é o emv de $\tau(\boldsymbol{\theta})$, $\tau(\boldsymbol{\theta}) = (\tau_1(\boldsymbol{\theta}), ..., \tau_k(\boldsymbol{\theta}))'$ e $\tau(\widehat{\boldsymbol{\theta}}) = (\tau_1(\widehat{\boldsymbol{\theta}}), ..., \tau_k(\widehat{\boldsymbol{\theta}}))'$.
- Precisamos particionar o espaço paramétrico associado à $\boldsymbol{\eta}$.
- Dem: Defina $L^*(\boldsymbol{\eta}) = \sup_{\boldsymbol{\theta}:\boldsymbol{\tau}(\boldsymbol{\theta})=\boldsymbol{\eta}} L(\boldsymbol{\theta}), \boldsymbol{\eta} \in \Lambda$, a verossimilhança de $\boldsymbol{\eta}$ induzida por $\boldsymbol{\tau}$.
- Seja $\Lambda = \boldsymbol{\tau}(\Theta)$ e para cada $\boldsymbol{\eta} \in \Lambda$, defina $\Theta_{\boldsymbol{\eta}} = \{\boldsymbol{\theta} \in \Theta : \boldsymbol{\tau}(\boldsymbol{\theta}) = \boldsymbol{\eta}\}$.
- Dessa forma, temos que $L^*(\boldsymbol{\eta}) = \sup_{\boldsymbol{\theta} \in \Theta_{\boldsymbol{\eta}}} L(\boldsymbol{\theta})$.

Teorema

- Além disso

$$L(\widetilde{\boldsymbol{\eta}}) = \sup_{\boldsymbol{\eta}\in\Lambda} L^*(\boldsymbol{\eta}) = \sup_{\boldsymbol{\eta}\in\Lambda} \sup\nolimits_{\boldsymbol{\theta}\in\Theta_{\boldsymbol{\eta}}} L(\boldsymbol{\theta}) = \sup_{\boldsymbol{\theta}\in\Theta} L(\boldsymbol{\theta}) = L(\widetilde{\boldsymbol{\theta}}) \tag{3}$$

- Entretanto, como $\widetilde{\boldsymbol{\theta}}$ é a emv de $\boldsymbol{\theta}$, vem que

$$L(\widetilde{\boldsymbol{\theta}}) = \sup\nolimits_{\boldsymbol{\theta}\in\Theta_{\widetilde{\boldsymbol{\eta}}}} L(\boldsymbol{\theta}) = L(\tau(\widetilde{\boldsymbol{\theta}})) \tag{4}$$

em que $\Theta_{\widetilde{\boldsymbol{\eta}}} = \left\{\widetilde{\boldsymbol{\theta}} \in \Theta : \boldsymbol{\tau}(\widetilde{\boldsymbol{\theta}}) = \widetilde{\boldsymbol{\eta}}\right\}$.

- Assim, de (3) e (4) temos que $L(\widetilde{\boldsymbol{\eta}}) = L(\boldsymbol{\tau}(\widetilde{\boldsymbol{\theta}}))$ e, assim, $\boldsymbol{\tau}(\widetilde{\boldsymbol{\theta}})$ é a emv de $\boldsymbol{\tau}(\boldsymbol{\theta})$.

Exemplo

- Exemplo: Seja $X_1, ..., X_n$ uma aa de $X \sim$ Bernoulli(θ) e $\tau(\theta) = \theta(1-\theta)$ (entra na variância exata de $\widehat{\theta}_{MV}$).
- Como $\widehat{\theta}_{MV} = \overline{X}$ então, pelo princípio da invariância $\widehat{\eta} = \overline{X}(1-\overline{X})$.
- Exemplo: Seja $X_i \stackrel{iid}{\sim} N(\mu, \sigma^2)$ e defina $\tau(\boldsymbol{\theta}) = P_{\boldsymbol{\theta}}(X \leq x_0)$, x_0 conhecido. Obtenha o emv de $\tau(\theta)$.
- Temos que

$$\eta = P_{\boldsymbol{\theta}}(X \leq x_0) = P_{\boldsymbol{\theta}}\left(Z \leq \frac{x_0 - \mu}{\sigma}\right)$$

Exemplo & Teorema

- Assim, como $\widehat{\mu} = \overline{X}$ e $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ são os emv de θ, então $\widehat{\eta} = P_{\boldsymbol{\theta}}\left(Z \leq \frac{x_0 - \widehat{\mu}}{\widehat{\sigma}}\right)$ e $\widehat{\sigma} = \sqrt{\widehat{\sigma}^2}$.
- Teorema: Seja $T = t(\boldsymbol{X})$ uma estatística suficiente para θ e $\widehat{\theta}$ seu respectivo emv. Então $\widehat{\theta} = f(t(\boldsymbol{X}))$ (vale também para o caso vetorial e/ou multivariado). Dem: exercício (usar o critério da fatoração).

Exemplo

- Exemplo: Seja $X_1, ..., X_n$ uma aa de X, $X \sim \exp(\theta)$. Obtenha o emv de x_q, o q-ésimo de X.
- Sabemos que:

$$F(x_q) = q, q \in (0,1),$$

mas,

$$1 - e^{-x_q/\theta} = q \rightarrow 1 - q = e^{-x_q/\theta} \rightarrow x_q = -\theta \ln(1-q).$$

- Assim, como $\widehat{\theta} = \overline{X}$ é o emv de θ, então $\widehat{x}_q = -\overline{X} \ln(1-q)$ é o emv de x_q.

Exemplo

- Seja $X_1, .., X_n$ uma a ade $X \sim U(\theta - 1/2, \theta + 1/2)$. Já provamos (págs. 33 e 34) que se $Y_n - \frac{1}{2} < \widehat{\theta} < Y_n + \frac{1}{2}$, $\widehat{\theta}$ é um emv de θ.
- Pode-se provar que

$$\widehat{\theta}_\gamma = Y_n - \frac{1}{2} + \gamma\left(1 + Y_1 - Y_n\right), \gamma \in (0,1)$$

- Sabemos que $(Y_1, Y_n)'$ é uma estatística suficiente para θ.
- Contudo, $\widehat{\theta}_\gamma = Y_n - \frac{1}{2} + \cos^2(X_1)\left(1 + Y_1 - Y_n\right)$ é um emv de θ, mas $\widehat{\theta}_\gamma \neq g(Y_1, Y_n)$ $(\widehat{\theta}_\gamma = g(Y_1, Y_n, X_1))$.

Exemplo

- Isto ocorre, essencialmente, pela fato de que o $\widehat{\theta}_{MV}$ não ser único.
- Exercício: Seja $X_1, ..., X_n$ uma aa de $X \sim \text{Pareto}(\alpha, \beta)$, ou seja,

$$f_X(x; \boldsymbol{\theta}) = \frac{\alpha \boldsymbol{\beta}^{\alpha}}{x^{\alpha+1}} \mathbb{1}_{[\beta,\infty)}(x).$$

Encontre o emv de $\boldsymbol{\theta} = (\alpha, \beta)' \in \mathcal{R}^+ \times \mathcal{R}^+$ e deduza todas as propriedades (esperança, distribuição etc) de forma exata, possíveis.

Método de mínimos quadrados (MMQ)

- Este é um dos métodos mais utilizados, quando não se quer fazer suposições acerca da distribuição dos dados (nem mesmo que pertença à família de localização-escala).
- Uma ampla apresentação pode ser encontrada aqui.
- Considere o modelo $Y_i = g_i(\boldsymbol{\theta}) + \xi_i$, $i = 1, 2, ..., n$, $\boldsymbol{\theta} \in \Theta \subseteq \mathcal{R}^k$ (sua suposição não é necessária para se considerar o método de MQ).
- Admita que as funções $g_i(.)$ são conhecidas e que $\xi_1, ..., \xi_n$ são erros não aleatórios e não correlacionados, tais que: $\mathcal{E}(\xi_i) = 0$ e $\mathcal{V}(\xi_i) = \sigma^2$ (em geral, desconhecido).
- Além disso, $g_i(.)$ é diferenciável em Θ, $\forall i$ (pelo menos duas vezes).

Método de mínimos quadrados (MMQ)

- A (uma) estimativa de MQ de $\boldsymbol{\theta}$ é obtida minimizando-se (através de algum procedimento apropriado) a expressão:

$$Q(\boldsymbol{\theta}) = \sum_{i=1}^{n} (y_i - g_i(\boldsymbol{\theta}))^2$$

em relação à $\boldsymbol{\theta}$. Portanto, se $\widetilde{\boldsymbol{\theta}}$ é a eqm de $\boldsymbol{\theta}$, então:

$$Q(\widetilde{\boldsymbol{\theta}}) = \sum_{i=1}^{n} (y_i - g_i(\widetilde{\boldsymbol{\theta}}))^2 = \inf_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^{n} (y_i - g_i(\boldsymbol{\theta}))^2.$$

- No caso unidimensional, devemos verificar se:

$$\left. \frac{\partial^2 Q(\theta)}{\partial \theta^2} \right|_{\theta = \widetilde{\theta}} > 0$$

Método de mínimos quadrados (MMQ)

- No caso bidimensional, devemos verificar se:

$$\left(\left.\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \theta_1^2}\right|_{\theta=\widetilde{\theta}} \right) \left(\left.\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \theta_2^2}\right|_{\theta=\widetilde{\theta}} \right) - \left(\left.\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2}\right|_{\theta=\widetilde{\theta}} \right)^2 > 0$$

e

$$\left.\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \theta_1^2}\right|_{\theta=\widetilde{\theta}} > 0 \text{ ou } \left.\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \theta_2^2}\right|_{\theta=\widetilde{\theta}} > 0$$

Método de mínimos quadrados (MMQ)

- No caso mais geral, devemos ter que $H(\boldsymbol{\theta})$

$$\boldsymbol{H}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \theta_1^2} & \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \theta_1 \theta_2} & \cdots & \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \theta_1 \theta_p} \\ \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \theta_2 \theta_1} & \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \theta_2^2} & \cdots & \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \theta_2 \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \theta_p \theta_1} & \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \theta_p \theta_2} & \cdots & \frac{\partial^2 Q(\boldsymbol{\theta})}{\partial \theta_p^2} \end{bmatrix}$$

deve ser positiva definida. Não, necessariamente, temos garantias de que o ponto é de mínimo.

Método de mínimos quadrados (MMQ)

- Podemos estimar σ^2 através de $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - g_i(\widehat{\theta}))^2$.
- Exemplo 1: Seja $Y_i = \theta + \xi_i$, $i = 1, 2, ..., n$. Nest caso, $g_i(\theta) = \theta$, $\forall i$, $\theta \in \Theta \subseteq \mathcal{R}$. Assim $h(\theta) = \sum_{i=1}^{n}(y_i - \theta)^2$
- Podemos provar que $\widehat{\theta} = \overline{Y}$ é o emq de θ (exercício). Se, além disso, $\xi \overset{iid}{\sim} N(0, \sigma^2)$, então $\widehat{\theta} = \overline{Y}$ também é o emv de θ.
- Exemplo 2: Seja $Y_i = \alpha + \beta x_i + \xi_i$, $i = 1, 2, ..., n$. Neste caso, $g_i(\theta) = \alpha + \beta x_i$, $\forall i$, $\boldsymbol{\theta} = (\alpha, \beta)' \in \Theta \subseteq \mathcal{R}^2$.
- Pode-se provar que a função: $h(\boldsymbol{\theta}) = \sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2$, atinge o mínimo em $\widetilde{\theta} = (\widetilde{\alpha}, \widetilde{\beta})'$, em que (exercício):

$$\widetilde{\alpha} = \overline{y} - \widetilde{\beta}\overline{x} \text{ e } \widetilde{\beta} = \frac{\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$

Método de mínimos quadrados (MMQ)

- Consequentemente $\widehat{\alpha} = \overline{Y} - \widehat{\beta}\overline{x}$ e $\widehat{\beta} = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})(X_i - \overline{x})}{\sum_{i=1}^{n}(X_i - \overline{x})^2}$, são os emq de $\boldsymbol{\theta}$.
- Neste caso, se $\xi_i \overset{iid}{\sim} N(0, \sigma^2)$, $\widehat{\boldsymbol{\theta}} = (\widehat{\alpha}, \widehat{\beta})'$ também é o emv de $\boldsymbol{\theta}$ (exercício).
- Em situações mais gerais, pode-se utilizar outras funções de perda (distâncias) $(g(\boldsymbol{\theta}))$, função dos valores observados e esperados.

Método de mínimos quadrados (MMQ)

- Nesses casos, a emq de $\boldsymbol{\theta}$ é obtido através da solução (das equações)

$$\sum_{i=1}^{n} d(y_i, \mathcal{E}(y_i; \widetilde{\boldsymbol{\theta}})) = \inf_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^{n} d(y_i, \mathcal{E}(y_i; \boldsymbol{\theta})),$$

em que $\mathcal{E}(Y_i; \boldsymbol{\theta}) = \mathcal{E}_{\boldsymbol{\theta}}(Y_i)$ e $d(.,.)$ é uma função de perda (distância) apropriada, por exemplo

$$d(y_i; \mathcal{E}(Y_i; \theta)) = |y_i - \mathcal{E}(Y_i; \boldsymbol{\theta})|.$$

- Veja o exemplo da regressão quantílica: aqui.

Multiplicadores de Lagrange

- As vezes, temos interesse em obter os estimadores de $\boldsymbol{\theta}$, sob alguma restrição ou, isto é mais conveniente.
- Por exemplo, poderíamos estar interessados em obter o emv de $\boldsymbol{\theta} \in \Theta \subseteq \mathcal{R}^k$, sujeito à $g_i(\boldsymbol{\theta}) = 0, \forall i = 1, 2..., r, r \leq k$ (não guarda, necessariamente, relação com a função definda para o método de MQ)
- Neste caso particular, podemos utilizar o métodos dos multiplicadores de Lagrande, ou seja, teremos que maximizar

$$g(\boldsymbol{\theta}, \boldsymbol{\lambda}) = l(\boldsymbol{\theta}) - \boldsymbol{\lambda}' g(\boldsymbol{\theta}) = l(\boldsymbol{\theta}) - \sum_{i=1}^{k} \lambda_i g_i(\boldsymbol{\theta})$$

$\boldsymbol{\lambda} = (\lambda_1, .., \lambda_k)' \in \mathcal{R}^r$.

Multiplicadores de Lagrange

- Exemplo, $\boldsymbol{X} \sim \text{multinomial}(n, \boldsymbol{\theta})$, n conhecido, $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)'$, $\boldsymbol{X} = (X_1, ..., X_k)'$, $\sum_{i=1}^{k} x_i = n$ e $\sum_{i=1}^{k} \theta_i = 1, \theta_i \in (0, 1)$ e $x_i \in \{0, 1, 2, ..., n\}$, $\forall i$. Encontre o emv de $\boldsymbol{\theta}$. Temos que

$$\frac{n!}{\prod_{i=1}^{k} x_i!} \prod_{i=1}^{k} \theta_i^{x_i} \mathbb{1}_A(\boldsymbol{x})$$

em que A é o conjunto que considera as restrições $\sum_{i=1}^{k} x_i = n, x_i \in \{0, 1, 2, ..., n\}$, veja também: aqui.

- Para obter o emv de $\boldsymbol{\theta}$, podemos maximizar

$$g(\boldsymbol{\theta}, \lambda) = l(\boldsymbol{\theta}) - \lambda \left(\sum_{i=1}^{k} \theta_i - 1 \right) = \sum_{i=1}^{k} x_i \ln(\theta_i) - \lambda \left(\sum_{i=1}^{k} \theta_i - 1 \right) + \text{const}$$

Multiplicadores de Lagrange

- Assim,

$$
\begin{aligned}
S(\theta_i) &= \frac{\partial g(\boldsymbol{\theta}, \lambda)}{\partial \theta_i} = \frac{x_i}{\theta_i} - \lambda, i = 1, 2, ..., k \\
S(\lambda) &= \sum_{i=1}^{k} \theta_i - 1
\end{aligned}
$$

- O sistema associado às equações de verossimilhança é dado por:

$$
\begin{cases} S(\widetilde{\theta}_i) = 0, i = 1, 2, ..., k \\ S(\widetilde{\lambda}) = 0 \end{cases} \rightarrow \begin{cases} \frac{x_i}{\widetilde{\theta}_i} - \widetilde{\lambda} = 0, i = 1, 2, ..., k(*) \\ \sum_{i=1}^{k} \widetilde{\theta}_i - 1 = 0(**) \end{cases} \rightarrow
$$

Multiplicadores de Lagrange

- De (**), temos que $\sum_{i=1}^{k} \widetilde{\theta}_i = 1(***)$. Portanto, de $(***)$ em (*), vem que

$$\sum_{i=1}^{k} x_i = \widetilde{\lambda} \sum_{i=1}^{k} \widetilde{\theta}_i \rightarrow \widetilde{\lambda} = \sum_{i=1}^{k} x_i = n(****)$$

- Finalmente, de (****) em cada equação de (*), temos que $\widetilde{\theta}_i = \frac{x_i}{n}$ e, assim, $\widehat{\theta}_i = \frac{X_i}{n}$ (calcular esperança, variância e eqm de $\widehat{\theta}_i$).
- Exercício: provar que a matriz Hessiana $H(\boldsymbol{\theta}, \lambda)$ é negativa definida e, portanto $\widehat{\theta}_i$ é o emv de θ_i, $i = 1, 2, ..., K$.
- Inferência Estatística restrita: aqui.

Método da substituição

- Seja $\widehat{\boldsymbol{\theta}}$ algum estimador (emv,emm,emq) de $\boldsymbol{\theta} \in \Theta \subseteq \mathcal{R}^k$ e $\boldsymbol{\tau}(\boldsymbol{\theta}) = (\tau_1(\boldsymbol{\theta}), ..., \tau_k(\boldsymbol{\theta}))'$ funções de interesse. Então, um estimador, digamos $\boldsymbol{\tau}(\widehat{\boldsymbol{\theta}})$ para $\boldsymbol{\tau}(\boldsymbol{\theta})$, é obtido através de

$$\boldsymbol{\tau}(\widehat{\boldsymbol{\theta}}) = (\tau_1(\widehat{\boldsymbol{\theta}}), ..., \tau_k(\widehat{\boldsymbol{\theta}}))'.$$

- Exemplo: Seja $X_1, ..., X_n \overset{iid}{\sim} U(0,\theta), \theta > 0$. Então $\widehat{\theta} = 2\overline{X}$ é o emm de θ. Seja $\tau(\theta) = \mathcal{V}(X) = \frac{\theta^2}{12}$.
- Logo um estimador para $\tau(\theta)$ (que não será, necessariamente, o emm dele) é dado por $\tau(\widehat{\theta}) = \frac{\widehat{\theta}^2}{12} = \frac{\overline{X}^2}{3}$.

Comparação de estimadores

- Já vimos como comparar estimadores, segundo alguns critérios como valor esperado, variância e EQM.
- Nesta parte aprofundaremos tal discussão.
- Suponha que queremos estimar θ, associado à $X \sim f_X(,;\theta)$ com base em uma amostra aleatória $X = (X_1, ..., X_n)'$.
- Como vimos, não é possível obtermos estimadores uniformemente ótimos ($\forall \theta \in \Theta$ e $\forall F_X(;\theta)$).
- Assim, para obtermos estimadores ótimos, deveremos nos restringer à classes específicas (sob certas restrições).

Comparação de estimadores

- Pergunta:
  1) Que propriedades um (bom) estimador deve ter?
  2) Que critérios devemos utilizar para selecionarmos um estimador ótimo?
- Exemplo: Considere dois estimadores, $\widehat{\theta}_1$ e $\widehat{\theta}_2$, de sorte que $\widehat{\theta}_1 \sim U(\theta - 1/2, \theta + 1/2)$ e $\widehat{\theta}_2 \sim U(\theta - 1, \theta + 1)$. Assim

$$\mathcal{E}(\widehat{\theta}_1) = \theta \quad ; \quad \mathcal{E}(\widehat{\theta}_2) = \theta$$
$$\mathcal{V}(\widehat{\theta}_1) = \frac{1}{12} \quad ; \quad \mathcal{V}(\widehat{\theta}_2) = \frac{1}{3}$$

- Nesse caso $\widehat{\theta}_1$ seria preferível à $\widehat{\theta}_2$, se utilizarmos a variância como critério de escolha (dado que os valores esperados são iguais).
- Em geral, busca-se $\mathcal{E}(\widehat{\theta}) \overset{\approx}{=} \theta$ (aproximadamente ou igual) e $\mathcal{V}(\widehat{\theta}) \approx 0$ (pequena).

Comparação de estimadores

- Def: Um estimador $\widehat{\theta}$ de θ (ou de $\tau(\theta)$) é dito ser não viciado (ou não viesado ou não tendencioso) se:

$$\mathcal{E}_\theta(\widehat{\theta}) = \theta \;\; [\mathcal{E}_\theta(\widehat{\theta}) = \tau(\theta)], \forall \theta \in \Theta \subseteq \mathcal{R}$$

- Exemplo 1: Seja $X_i \overset{iid}{\sim}$ Bernoulli(θ), então $\widehat{\theta} = \overline{X}$ é um estimador não viciado de θ.
- Exemplo 2: Seja $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$. Nesse caso, $\widehat{\boldsymbol{\theta}} = \left(\overline{X}, \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right)'$ é viciado para $\boldsymbol{\theta}$ pois, apesar de $\overline{X}$ ser não viciado para μ, $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ é viciado para σ^2.

Comparação de estimadores

- Exemplo 3: Seja $X_1, ..., X_n \overset{iid}{\sim}$ Poisson(θ), $\theta > 0$. Seja $\widehat{\tau}(\theta) = (-2)^{X_1}$ um estimador para $\tau(\theta) = e^{-3\theta}$. Note que

$$\mathcal{E}_\theta(\widehat{\tau}(\theta)) = \sum_{x=0}^{\infty}(-2)^x \frac{e^{-\theta}\theta^x}{x!} = e^{-\theta}\sum_{x=0}^{\infty}\frac{(-2\theta)^x}{x!} = e^{-3\theta} = \tau(\theta)$$

  Contudo, apesar de $\widehat{\tau}(\theta)$ ser não viciado, note que $P(\widehat{\tau}(\theta) < 0) > 0$. Assim, tal estimador poderia não ser apropriado.

- Exemplo 4: Seja $X_1, ..., X_n$ uma aa de $X \sim$ Bernoulli(θ), $\theta \in (0,1)$ e $\tau(\theta) = \frac{\theta}{1-\theta}$. Neste caso, $\nexists$ env (estimador não viciado) de $\tau(\theta)$. Suponha, então, que $\exists$ tal estimador, então (lembrando que temos um total de 2^n amostras possíveis), temos que (slide seguinte):

Comparação de estimadores

- Assim

$$
\begin{aligned}
\mathcal{E}_{\theta}(\widehat{\theta}) &= \sum_{\boldsymbol{x}} \widetilde{\theta}(\boldsymbol{x})\theta^{n\overline{x}}(1-\theta)^{n(1-\overline{x})} = \frac{\theta}{1-\theta} \\
&\rightarrow \sum_{j=1}^{2^n} \widetilde{\theta}_j \theta^{k_j}(1-\theta)^{n-k_j} = \frac{\theta}{1-\theta} \\
&\rightarrow \sum_{j=1}^{2^n} \widetilde{\theta}_j \theta^{k_j-1}(1-\theta)^{n-k_j} = (1-\theta)^{-1}, \forall \theta \in \Theta
\end{aligned}
$$

O que é uma contradição pois está-se afirmando que um polinômio de grau n (em função de θ) é igual à $(1-\theta)^{-1}$. Assim, não existe um env para $\frac{\theta}{1-\theta}$, neste caso.

Comparação de estimadores

- Erro-quadrático médio (já fora definido). De uma forma geral, temos que $\mathcal{EQM}_{\theta}(\theta) = \mathcal{E}_{\theta}[(\widehat{\theta} - \theta)^2]$.
- Note que

$$
\begin{aligned}
\mathcal{E}_{\theta}[(\widehat{\theta} - \theta)^2] &= \mathcal{E}_{\theta}[(\widehat{\theta} - \mathcal{E}(\widehat{\theta}) - (\theta - \mathcal{E}(\widehat{\theta})))^2] \\
&= \mathcal{E}_{\theta}[(\widehat{\theta} - \mathcal{E}(\widehat{\theta}))^2] + \mathcal{E}_{\theta}[(\theta - \mathcal{E}(\widehat{\theta}))^2] \\
&- 2\mathcal{E}_{\theta}[(\widehat{\theta} - \mathcal{E}(\widehat{\theta}))(\theta - \mathcal{E}(\widehat{\theta}))] \\
&= \mathcal{V}_{\theta}(\widehat{\theta}) + \mathcal{B}^2(\widehat{\theta}) - 2(\theta - \mathcal{E}(\widehat{\theta}))\mathcal{E}_{\theta}[(\widehat{\theta} - \mathcal{E}(\widehat{\theta})] \\
&= \mathcal{V}_{\theta}(\widehat{\theta}) + \mathcal{B}^2(\widehat{\theta})
\end{aligned}
$$

A quantidade $\mathcal{B}_{\theta}(\widehat{\theta})$ é chamada de vício do estimador, em que se $\mathcal{B}_{\theta}(\widehat{\theta}) = 0 \rightarrow \mathcal{V}(\widehat{\theta}) = \mathcal{EQM}(\widehat{\theta})$.

Exemplo

- Seja $X_1, ..., X_n$ uma aa de $X \sim \text{Bernoulli}(\theta)$, $\theta \in (0,1)$
- Sejam $\widehat{\theta}_1$ e $\widehat{\theta}_2$, dois estimadores para θ, tais que $\widehat{\theta}_1 = \overline{X}$ e $\widehat{\theta}_2 = \frac{\alpha + n\overline{X}}{\alpha+\beta+n}, \alpha, \beta \in \mathcal{R}^+$.
- Pode-se provar (para $\widehat{\theta}_1$ já fora provado) que:

$$\mathcal{E}_\theta(\widehat{\theta}_1) = \theta \quad ; \quad \mathcal{E}_\theta(\widehat{\theta}_2) = \frac{\alpha + n\theta}{\alpha+\beta+n}$$

$$\mathcal{B}_\theta(\widehat{\theta}_1) = 0 \quad ; \quad \mathcal{B}_\theta(\widehat{\theta}_2) = \frac{\alpha + n\theta}{\alpha+\beta+n} - \theta$$

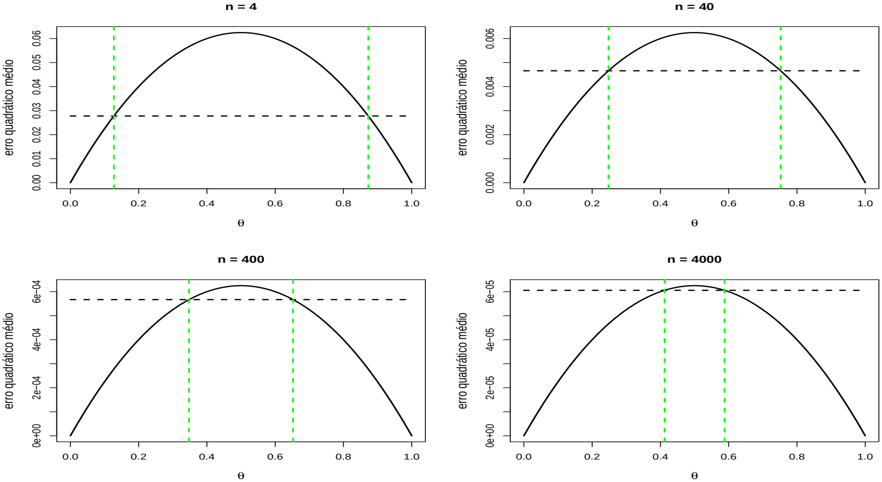
$$\mathcal{V}_\theta(\widehat{\theta}_1) = \frac{\theta(1-\theta)}{n} \quad ; \quad \mathcal{V}_\theta(\widehat{\theta}_2) = \frac{n\theta(1-\theta)}{(\alpha+\boldsymbol{\beta}+n)^2}$$

$$\mathcal{EQM}(\widehat{\theta}_1) = \frac{\theta(1-\theta)}{n} \quad ; \quad \mathcal{EQM}(\widehat{\theta}_2) = \frac{n\theta(1-\theta)}{(\alpha+\beta+n)^2} + \left[\frac{\alpha + n\theta}{\alpha+\beta+n} - \theta\right]^2$$

Exemplo

- Tomando $\alpha = \beta = \frac{\sqrt{n}}{2}$, temos que $\mathcal{EQM}(\widehat{\theta}_2) = \frac{n}{4(\sqrt{n}+n)^2}$.
- Graficamente (slides seguinte): linha sólida em negrito - $\widehat{\theta}_1$; linha sólida em negrito - $\widehat{\theta}_2$.
- Exercício: obter os pontos $\{a_n, b_n\}$, em relação à θ, tais que $\mathcal{EQM}(\widehat{\theta}_1) = \mathcal{EQM}(\widehat{\theta}_2)$ $(\alpha = \beta = \frac{\sqrt{n}}{2})$.
- Assim, em termos do $\mathcal{EQM}$, temos que o $\widehat{\theta}_1$ é melhor que $\widehat{\theta}_2$ para $\theta \in (0, a_n) \cup (b_n, 1)$

Comparação dos EQM em função de θ e n



Definições

- Se $\widehat{\theta}_1$ e $\widehat{\theta}_2$ são dois estimadores de θ, dizemos que $\widehat{\theta}_1$ é melhor do que $\widehat{\theta}_2$, em termos do EQM se

$$\mathcal{EQM}(\widehat{\theta}_1) \leq \mathcal{EQM}(\widehat{\theta}_2), \forall \theta \in \Theta.$$

- Se ambos forem não viciados, isso equivale a dizer que $\widehat{\theta}_1$ é melhor do que $\widehat{\theta}_2$, em termos da variância, ou seja

$$\mathcal{V}(\widehat{\theta}_1) \leq \mathcal{V}(\widehat{\theta}_2), \forall \theta \in \Theta.$$

- Embora não seja a forma mais apropriada, também podemos comparar os estimadores através das variâncias mesmo se pelo menos um deles for não viciado.

Estimadores ótimos

- Seja $C_\tau = \left\{\widehat{\theta} : \mathcal{E}(\widehat{\theta}) = \tau(\theta), \forall\theta \in \Theta\right\}$ a classe dos estimadores não viciados de $\tau(\theta)$ e vamos supor que $C_\tau \neq \emptyset$. Nosso objetivo é encontrar $\widehat{\theta}^*$ em C_τ,

$$\mathcal{V}(\widehat{\theta}^*) \leq \mathcal{V}(\widehat{\theta}), \forall\theta \in \Theta, \forall\, \widehat{\theta}^* \neq \widehat{\theta} \tag{5}$$

- Definição: um estimador pertencente à C_τ que satisfaz a (5) é dito ser um estimador não viciado de variância uniformemente mínima (ENVVUM).
- Na busca por esse tipo de estimador, os conceitos de redução de dados ser-nos-ão bem úteis.

Condições de regularidade

- Seja $\boldsymbol{X} = (X_1, ..., X_n)'$ uma aa de $X \sim f_X(.;\theta)$, $\theta \in \Theta \subseteq \mathcal{R}$. As condições de regularidade são:
  1) $I(\theta) < \infty, \forall \theta \in \Theta$, em que $I(\theta) = \mathcal{E}\left[\left(\frac{\partial}{\partial\theta}\ln f(\boldsymbol{X};\theta)\right)^2\right]$ (Informação de Fisher)
  2) $\forall T = t(\boldsymbol{X}), \mathcal{E}\left[|t(\boldsymbol{X})|\right] < +\infty$, então
  $$\frac{\partial}{\partial\theta}\int_{\mathcal{R}^n} t(\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x};\theta)d\boldsymbol{x} \text{ e } \int_{\mathcal{R}^n}\frac{\partial}{\partial\theta}t(\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x};\theta)d\boldsymbol{x}$$
  $\exists$ e são contínuas em θ
  3)
  $$\frac{\partial}{\partial\theta}\int_{\mathcal{R}^n} t(\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x};\theta)d\boldsymbol{x} = \int_{\mathcal{R}^n}\frac{\partial}{\partial\theta}t(\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x};\theta)d\boldsymbol{x}$$

Teorema de Cramer-Rao

- Seja $\boldsymbol{X} = (X_1, ..., X_n)'$ uma aa de $f_X(.;\theta), \theta \in \Theta$ e $T = t(\boldsymbol{X})$ um estimador não viciado de $\tau(\theta)$, $\tau : \Theta \to \mathcal{R}$ é uma função diferenciável. Se as condições de regularidade forem satisfeitas então:

$$\mathcal{V}(T) \geq \frac{\left[\frac{d}{d\theta}\tau(\theta)\right]^2}{I(\theta)}, \forall\theta \in \Theta \tag{6}$$

- Obs: As distribuições pertencentes à FE satisfazem as CR (denote: $\frac{d}{d\theta}\tau(\theta) = \tau'(\theta)$).
- A inequação (6) é conhecida como limite inferior de Cramer-Rao (LICR).
- Assim, todo e qualquer estimador não viciado cuja variância atingir o LICR será um ENVVUM.

Teorema de Cramer-Rao

- Demonstração: Sejam Y e W va's, $\mathcal{E}(Y) = 0$. Assim

$$\text{Cov}^2(Y, W) \leq \mathcal{V}(Y)\mathcal{V}(W) = \mathcal{E}(Y^2)\mathcal{V}(W)$$

- Portanto,

$$\text{Cov}^2\left(t(\boldsymbol{X}), \frac{\partial}{\partial\theta}\ln f_{\boldsymbol{X}}(\boldsymbol{X};\theta)\right) \leq \mathcal{V}(t(\boldsymbol{X}))\mathcal{E}\left[\left(\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{X}}(\boldsymbol{X};\theta)\right)^2\right] \quad (7)$$

Teorema de Cramer-Rao

- Mas

$$\begin{aligned}
\text{Cov}\left(t(\boldsymbol{X}), \frac{\partial}{\partial\theta}\ln f_{\boldsymbol{X}}(\boldsymbol{X};\theta)\right) &= \int_{\mathcal{R}^n} t(\boldsymbol{x})\left(\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{X}}(\boldsymbol{x};\theta)\right) f_{\boldsymbol{X}}(\boldsymbol{x};\theta)d\boldsymbol{x} \\
&= \int_{\mathcal{R}^n} t(\boldsymbol{x})\frac{\frac{\partial}{\partial\theta}f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{f_{\boldsymbol{X}}(\boldsymbol{x};\theta)} f_{\boldsymbol{X}}(\boldsymbol{x};\theta)d\boldsymbol{x} \\
&= \frac{\partial}{\partial\theta}\int_{\mathcal{R}^n} t(\boldsymbol{x})f_{\boldsymbol{X}}(\boldsymbol{x};\theta)d\boldsymbol{x} = \frac{\partial}{\partial\theta}\mathcal{E}(t(\boldsymbol{X})) \\
&= \tau'(\theta) \qquad (8)
\end{aligned}$$

- Logo, de (8) em (7) vem que

$$\mathcal{V}(T) \geq \frac{\left[\frac{d}{\partial\theta}\tau(\theta)\right]^2}{I(\theta)}, \forall\theta \in \Theta$$

Teorema de Cramer-Rao

- Exercício. Prove, sob as condições de regularidade, que:

$$\mathcal{E}\left\{\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{X}}(\boldsymbol{X};\theta)\right\} = 0; \text{ e } I(\theta) = \mathcal{E}\left\{-\frac{\partial^2}{\partial\theta^2}\ln f_{\boldsymbol{X}}(\boldsymbol{X};\theta)\right\}$$

- OBS: Se $X_1, ..., X_n$ é uma aa de $f_X(.;\theta)$ então $I(\theta)$ (associada a $f_{\boldsymbol{X}}(.;\theta)$) é tal que: $I(\theta) = nI_1(\theta)$, em que

$$I_1(\theta) = \mathcal{E}\left\{-\frac{\partial^2}{\partial\theta^2}\ln f_X(X;\theta)\right\}$$

Teorema de Cramer-Rao

- Corolário: Sob as condições do teorema,

$$\mathcal{V}_\theta(T) = \frac{\left[\frac{d}{d\theta}\tau(\theta)\right]^2}{I(\theta)} \leftrightarrow \exists a(\theta), a(\theta)[t(\boldsymbol{x}) - \tau(\theta)] = \frac{\partial}{\partial\theta} \ln f_{\boldsymbol{X}}(\boldsymbol{x};\theta), \forall\theta \in \Theta$$

- Prova: Note que

$$Cov\left(T, \frac{\partial}{\partial\theta} \ln f_{\boldsymbol{X}}(\boldsymbol{x};\theta)\right) = Cov\left(T - \tau(\theta), \frac{\partial}{\partial\theta} \ln f_{\boldsymbol{X}}(\boldsymbol{x};\theta)\right)$$

Teorema de Cramer-Rao

- A igualdade (no LICR) ocorre se, e somente se,

$$Cov(Y, W) = E(Y^2)\mathcal{V}(W) \leftrightarrow Corre(Y, W) = 1,$$

ou seja se $Y = aW + b$, para algum vetor (a,b) (não aleatório). Ou seja, se e somente se, $\exists$ constantes $a^*(\theta)$ e $b^*(\theta)$

$$T - \tau(\theta) = a^*(\theta)\frac{\partial}{\partial\theta}\ln f_{\boldsymbol{X}}(\boldsymbol{x};\theta) + b^*(\theta)$$

- Cont. ou seja, se:

$$\frac{T - \tau(\theta)}{a^*(\theta)} = \frac{\partial}{\partial\theta}\ln f_{\boldsymbol{X}}(\boldsymbol{x};\theta) + \frac{b^*(\theta)}{a^*(\theta)}$$

- Assim, se $a^*(\theta) = a(\theta)^{-1}$ e $b^*(\theta) = 0$, o $LICR(\tau(\theta))$ é atingido.

LICR

- Exemplo 1: Seja $X_1, ..., X_n$ uma aa de $X \sim \exp(\theta)$, $\mathcal{E}(X) = \theta$. Encontre o $LICR(\theta)$.
- Temos que:

$$
\begin{aligned}
H(\theta) &= \frac{\partial^2}{\partial\theta^2} \ln f_{\boldsymbol{X}}(\boldsymbol{x};\theta) = \frac{n}{\theta^2} - \frac{2n\overline{x}}{\theta^3} \\
I(\theta) &= \mathcal{E}(-H(\theta)) = \frac{-n}{\theta^2} + \frac{2n}{\theta^2} = \frac{n}{\theta^2}
\end{aligned}
$$

- Então o $LICR(\theta) = \frac{\theta^2}{n}$.
- Por outro lado, sabemos que $\widehat{\theta}_{MV} = \overline{X}$, de sorte que, $\mathcal{E}(\widehat{\theta}_{MV}) = \theta$ e $\mathcal{V}(\widehat{\theta}_{MV}) = \frac{\theta^2}{n}$. Assim, neste caso, $\widehat{\theta}_{MV}$ é um ENVVUM para θ.

LICR

- Exemplo 1.1: Vamos agora encontrar o $LICR(\theta) = \frac{1}{\theta}$. Temos que $LICR(\tau(\theta)) = \frac{[\tau'(\theta)]^2}{I(\theta)} = \frac{1}{n\theta^2}$.
- Vamos estudar o comportamento do estimador $\widehat{\theta} = \frac{1}{\overline{X}} = \frac{n}{S}$, $S = \sum_{i=1}^{n} X_i$. Note que $S \sim \text{gama}(n,\theta)$. Assim:

$$
\begin{aligned}
\mathcal{E}(\widehat{\theta}^r) &= \int_0^\infty \frac{1}{\theta^n\Gamma(n)}\frac{n^r}{s^r}e^{-s/\theta}s^{n-1}ds \\
&= \frac{n^r}{\theta^n\Gamma(n)}\theta^{n-r}\Gamma(n-r)\underbrace{\int_0^\infty \frac{1}{\theta^{n-r}\Gamma(n-r)}e^{-s/\theta}s^{n-r-1}ds}_{1} \\
&= \frac{n^r}{\theta^r}\frac{\Gamma(n-r)}{\Gamma(n)}
\end{aligned}
$$

LICR

- Assim, $\mathcal{E}(\widehat{\theta}) = \frac{n}{\theta}\frac{1}{n-1} \to \mathcal{E}\left(\frac{n-1}{n}\widehat{\theta}\right) = \frac{1}{\theta} \to \mathcal{E}(\widehat{\theta}^*) = \frac{1}{\theta}$, $\widehat{\theta}^* = \frac{n-1}{n}\widehat{\theta}$
- Além disso, $\mathcal{E}(\widehat{\theta}^2) = \frac{n^2}{\theta^2(n-1)(n-2)}$. Assim,

$$\mathcal{V}(\widehat{\theta}) = \frac{n^2}{\theta^2(n-1)(n-2)} - \frac{n^2}{\theta^2(n-1)^2} = \frac{n^2}{\theta^2(n-1)^2(n-2)}$$

- Portanto, $\mathcal{V}(\widehat{\theta}^*) = \frac{(n-1)^2}{n^2}\mathcal{V}(\widehat{\theta}) = \frac{1}{\theta^2(n-2)} > \frac{1}{\theta^2 n}$. Assim, não podemos afirmar se $\widehat{\theta}^*$ é ou não um ENVVUM para $\frac{1}{\theta}$.
- Exercício: para este modelo, encontrar $LICR(\tau(\theta))$, em que $\tau(\theta) = \theta^2$ e $\tau(\theta) = \frac{1}{\theta^2}$

LICR

- Exemplo: Seja $X_1, ..., X_n$ uma aa de $X \sim N(\mu, \sigma^2)$, com μ conhecido. Achar o $LICR(\sigma^2)$. Já vimos que:

$$
\begin{aligned}
S(\sigma^2) &= \frac{\partial}{\partial \sigma^2} \ln f_{\boldsymbol{X}}(\boldsymbol{x}; \sigma^2) = -\frac{n}{\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (x_i - \mu)^2 \\
&= \frac{n}{2(\sigma^2)^2} \left[\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 - \sigma^2 \right] \\
&= a(\sigma^2)[t(\boldsymbol{x}) - \tau(\sigma^2)] \qquad (9)
\end{aligned}
$$

Como $\mathcal{E}[t(\boldsymbol{X})] = \mathcal{E}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right) = \sigma^2$ e por (9), temos que $\mathcal{V}(t(\boldsymbol{X})) = LICR(\sigma^2)$ e, assim, $t(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ é um ENVVUM de σ^2, pelo colorário.

LICR

- Exercício: Seja $X_1, ..., X_n$ uma aa de $X \sim$ Poisson(θ). Prove que se $\widehat{\theta} = \overline{X}$, então $\mathcal{V}(\widehat{\theta}) = LICR(\theta)$.
- Exemplo: Considere o modelo do exercício anterior e $\tau(\theta) = e^{-\theta}(1 + \theta)$.
- Note que $P_\theta(X = 0) = e^{-\theta}$ e $P_\theta(X = 1) = e^{-\theta}\theta$.
- Defina $T = I_{\{X=0 \cup X=1\}}$. Assim, $\mathcal{E}(T) = P_\theta(X = 0 \cup X = 1) = e^{-\theta} + e^{-\theta}\theta = e^{-\theta}(1 + \theta) = \tau(\theta)$.
- Seja agora $T^* = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i=0 \cup X_i=1\}}$, assim $\mathcal{E}(T^*) = \tau(\theta)$ e $\mathcal{V}(T^*) = \frac{\tau(\theta)(1-\tau(\theta))}{n}$
- Por outro lado, vamos verificar se algum estimador não viciado de $\tau(\theta)$, neste caso, atinge o LICR.

LICR

- Temos que $S(\theta) = \frac{n}{\theta}[\overline{x} - \theta]$
- Se T é um estimador não viesado de $\tau(\theta)$ cuja $\mathcal{V}_{\theta}(T) = LICR(\tau(\theta))$, então

$$
\begin{aligned}
a(\theta)[t(\boldsymbol{x}) - \tau(\theta)] &= \frac{\partial}{\partial\theta} \ln f_{\boldsymbol{X}}(\boldsymbol{x};\theta) \\
& t(\boldsymbol{x}) = \frac{n}{\theta a(\theta)}[\overline{x} - \theta] + \tau(\theta) \\
&= \frac{n\overline{x}}{\theta a(\theta)} + \left[\tau(\theta) - \frac{n}{a(\theta)}\right]
\end{aligned}
$$

LICR

- Mas para que $t(\boldsymbol{x})$ seja uma estatística, devemos ter

$$\tau(\theta) - \frac{n}{a(\theta)} = 0 \rightarrow a(\theta) = \frac{\tau(\theta)}{n} \rightarrow t(\boldsymbol{x}) = \frac{n^2\overline{x}}{\tau(\theta)}$$

o que é uma contradição. Logo,

$$\mathcal{V}(T) > LICR(\tau(\theta)), \forall\theta \in \Theta, \forall T, \mathcal{E}_{\theta}(T) = \tau(\theta)$$

Teorema de Rao-Blackwell

- Exemplo: Seja $X_1, ..., X_n$ uma aa de $X \sim \text{Bernoulli}(\theta), \theta \in (0,1)$. Defina $T = X_1$, $S = \sum_{i=1}^n X_i$ e $T^* = \mathcal{E}(T|S)$. Note que

$$
\begin{aligned}
\mathcal{E}(T|S=s) &= P(X_1 = 1|S=s) = \frac{P(X_1 = 1, S = s)}{P(S=s)} = \\
&= \frac{P(X_1 = 1, \sum_{i=1}^n X_i = s)}{\binom{n}{s}\theta^s(1-\theta)^{n-s}} = \frac{P(X_1 = 1, \sum_{i=2}^n X_i = s-1)}{\binom{n}{s}\theta^s(1-\theta)^{n-s}} \\
&= \frac{\binom{n-1}{s-1}\theta^{s-1+1}(1-\theta)^{n-1-s+1}}{\binom{n}{s}\theta^s(1-\theta)^{n-s}} = \frac{s}{n} = \frac{\sum_{i=1}^n x_i}{n}
\end{aligned}
$$

Note que $\mathcal{V}(X_1) = \theta(1-\theta) \geq \mathcal{V}(T^*) = \frac{\theta(1-\theta)}{n}, \forall \theta \in \Theta$.

Teorema de Rao-Blackwell

- Seja $T = t(\boldsymbol{X})$ um estimador não viciado para $\tau(\theta), \forall \theta \in \Theta$ e $S = s(\boldsymbol{X})$ uma estatística suficiente para $\tau(\theta)$. Defina $T^* = \mathcal{E}(T|S)$. Então:
  1) T^* é uma estatística.
  2) $\mathcal{E}(T^*) = \tau(\theta), \forall \theta \in \Theta$.
  3) $\mathcal{V}(T^*) \leq \mathcal{V}(T), \forall \theta \in \Theta$.
- Prova: 1) Segue do fato de que S é suficiente e $T = t(\boldsymbol{X})$.
- Prova: 2) $\mathcal{E}(T^*) = \mathcal{E}(\mathcal{E}(T|S)) = \mathcal{E}(T) = \tau(\theta)$, pois T é um env de $\tau(\theta)$.

Teorema de Rao-Blackwell

- Prova: 3) Temos que

$$\begin{aligned}
\mathcal{V}(T^*) &= \mathcal{V}(\mathcal{E}(T|S)) \\
\mathcal{V}(T) &= \mathcal{E}(\mathcal{V}(T|S)) + \mathcal{V}(\mathcal{E}(T|S)) \\
&= \underbrace{\mathcal{E}(\mathcal{V}(T|S))}_{\geq 0} + \mathcal{V}(T^*) \\
&\rightarrow \mathcal{V}(T) \geq \mathcal{V}(T^*)
\end{aligned}$$

- Exercício: Seja $X_1, ..., X_n$ uma aa de $X \sim$ Poisson(θ) e $\tau(\theta) = e^{-\theta}(1+\theta)$. Obtenha, explicitamente, $T^{**} = \mathcal{E}(T|S)$ em que $T^* = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i=0 \cup X_i=1\}}$ e $S = \sum_{i=1}^{n} X_i$. Mostre que: $\mathcal{V}(T^{**}) \leq \mathcal{V}(T^*)$.

Teorema de Rao-Blackwell (TRB)

- Observação: Se T é um estimador de $\tau(\theta)$ (não necessariamente não viciado) e S é uma estatística suficiente, então $\mathcal{E}_\theta[(T^* - \tau(\theta))^2] \leq \mathcal{E}_\theta[(T - \tau(\theta))^2], \forall \theta \in \Theta$ (EQM), em que $T^* = \mathcal{E}(T|S)$. Sugestão para demonstração: $\mathcal{E}(T) = \mathcal{E}(T^*)$.
- Exemplo (necessidade da Suficiência de S): Seja X_1, X_2 uma aa de $X \sim N(\mu, \sigma^2)$, da tamanho 2 e defina $T = \frac{X_1+X_2}{2}$. Note que:

$$
\begin{aligned}
\mathcal{E}(T|X_1) &= \mathcal{E}\left[\left(\frac{X_1+X_2}{2}\middle| X_1\right)\right] = \frac{1}{2}\mathcal{E}\left[\mathcal{E}(X_1|X_1) + \mathcal{E}(X_2|X_1)\right] \\
&= \frac{1}{2}(X_1 + \mu)
\end{aligned}
$$

que não é uma estatística $(\mathcal{E}(T|X_1))$.

Teorema

- Seja T um ENVVUM de $\tau(\theta)$ ($\mathcal{V}_\theta(T) \leq \infty$). Então T é único.
- Dem: Seja T' um outro ENVVUM de $\tau(\theta)$ e defina $T^* = \frac{T+T'}{2}$. Assim

$$\mathcal{E}(T^*) = \tau(\theta)$$

e

$$\begin{aligned}
\mathcal{V}(T^*) &= \frac{1}{4}\mathcal{V}(T) + \frac{1}{4}\mathcal{V}(T') + \frac{1}{2}Cov(T, T') \\
&\underbrace{\leq}_{\text{Cauchy-Schwarz}} \frac{1}{4}\mathcal{V}(T) + \frac{1}{4}\mathcal{V}(T') + \frac{1}{2}\left(\mathcal{V}(T)\mathcal{V}(T')\right)^{1/2} \\
&\underbrace{=}_{\mathcal{V}(T)=\mathcal{V}(T')} \mathcal{V}(T) \rightarrow \mathcal{V}(T^*) \leq \mathcal{V}(T), \forall \theta \in \Theta
\end{aligned}$$

Teorema

- No entanto, como T é ENVVUM, $\mathcal{V}(T^*) = \mathcal{V}(T), \forall\theta \in \Theta$. Por outro lado, isso implica que:

$$
\begin{aligned}
& \frac{1}{4}\mathcal{V}(T) + \frac{1}{4}\mathcal{V}(T') + Cov(T, T') = \\
& \frac{1}{4}\mathcal{V}(T) + \frac{1}{4}\mathcal{V}(T') + (\mathcal{V}(T)\mathcal{V}(T'))^{1/2} \\
Cov(T, T') \quad = \quad & (\mathcal{V}(T)\mathcal{V}(T'))^{1/2} (= \mathcal{V}(T)) \\
\rightarrow \quad & Cov(T, T') = \pm 1 \rightarrow \exists\, a(\theta), h(\theta) \text{ constantes}, \\
& T' = a(\theta)T + h(\theta) \rightarrow \mathcal{E}(T') = a(\theta)\tau(\theta) + h(\theta) \text{ e} \\
& \mathcal{V}(T') = a^2(\theta)\mathcal{V}(T)
\end{aligned}
$$

Assim, como T' também é um ENVVUM, temos que $a(\theta) = 1$ e $h(\theta) = 0$, $\forall\theta \in \Theta$. Logo $T = T'$.

Teorema de Lehmann-Scheffe

- Já vimos que, se existir, o ENVVUM é único.
- O LICR não, necessariamente, é útil, para obter o ENVVUM.
- O TRB apenas, em princípio, aperfeiçoa estimadores não viciados, no sentido de diminuir sua variâncias.
- O Teorema de Lehmann-Scheffe (TLS), sob certas condições, produz o ENVVUM.

Teorema de Lehmann-Scheffe

- Seja S uma estatística suficiente e completa e T um estimador não viciado de $\tau(\theta)$, tal que, $T = g(S)$. Então T é o ENVVUM.
- Prova:
  1) Unicidade: Seja T' um outro ENVVUM, assim $T' = h(S)$ (pelo TRB). Então, defina $f(S) = T - T' = g(S) - h(S)$. Portanto

     $$\text{Se}, \mathcal{E}(f(S)) = 0 \rightarrow f(S) = 0, \forall a \in B_s \,(\text{suporte de S}), \text{pois S é completa}$$

     $$\text{Logo, } T - T' \equiv 0 \rightarrow T \equiv T', \forall, \theta \in \Theta.$$

  2) Existência: Seja T' um estimador não viciado de $\tau(\theta)$. Então $T^* = \mathcal{E}(T|S)$ é um estimador não viciado de $\tau(\theta)$ e $\mathcal{V}_\theta(T^*) \leq \mathcal{V}_\theta(T)$, $\forall \theta \in \Theta$. Como T^* é uma função de S, pelo item 1), temos que $T^* \equiv T$. Assim, $\mathcal{V}_\theta(T^*) = \mathcal{V}_\theta(T), \forall \theta \in \Theta$.

Teorema de Lehmann-Scheffe

- Exemplo: Seja $X_1, ..., X_n$ uma aa de $X \sim \exp(\theta)$, $\mathcal{E}(X) = \theta$ e defina $\tau(\theta) = \frac{1}{\theta}$. Já vimos que, $\widehat{\theta} = \frac{n-1}{n}\frac{1}{S}$ é uma estimador não viciado de $\tau(\theta)$.
- Como S é uma estatística suficiente e completa, $\widehat{\theta} = g(S)$ e $\mathcal{E}(\widehat{\theta}) = \tau(\theta)$, pelo TLS, temos que $\widehat{\theta}$ é o ENVVUM de $\tau(\theta)$.
- Obs: Isso é verdade, ainda que

$$\mathcal{V}_\theta(\widehat{\theta}) > LICR(\tau(\theta)), \forall \theta \in \Theta, \forall n \in \mathcal{N}.$$

o que, de fato, acontece neste caso.

Exemplo

- Seja $X_1, ..., X_n$ uma aa de $X \sim \text{Poisson}(\theta)$ e $\tau(\theta) = e^{-\theta}(1+\theta)$. Obtenha o ENVVUM de τ.
- Já vimos (fora deixado para ser provado) que

$$T^{**} = \mathcal{E}(T^*|S) = \left(\frac{n-1}{n}\right)^S \left(1 + \frac{S}{n-1}\right),$$

em que $T^* = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i=0 \cup X_i=1\}}$ e $S = \sum_{i=1}^n X_i$.

Exemplo

- Logo, T^{**} é o ENVVUM de $\tau(\theta)$, pelo TLS, pois $\mathcal{E}(T^{**}) = \mathcal{E}(T^{*}) = \tau(\theta)$ (como já vimos anteriormente) e $T^{**} = g(S)$, S uma estatística suficiente e completa.
- Exemplo: Seja $X_1, .., X_n$ uma aa de $X \sim U(0,\theta)$, $\theta > 0$. O ENVVUM de θ é $\widehat{\theta} = \frac{n+1}{n} Y_n$, $Y_n = \max\{X_1, ..., X_n\}$. Pelo TLS, como $\mathcal{E}(\widehat{\theta}) = \theta$ e $\widehat{\theta} = g(Y_n)$, em que Y_n é uma estatística suficiente e completa.

Formas de obter o ENVVUM ($\tau(\theta)$)

1) Obter uma estatística suficiente e completa (S) e
   1.1) Construir um env, digamos, T, que seja função de S ($T = g(S)$), tal que $\mathcal{E}(T) = \tau(\theta)$ (TLS).
   1.2) Obter $T = \mathcal{E}(T'|S)$, em que T' é um unv de $\tau(\theta)$ (TLS + TRB).
2) Calcular o LICR e provar que $\mathcal{V}(T) = LICS(\tau(\theta))$, $\forall\theta \in \Theta$, $\mathcal{E}(T) = \tau(\theta)$ (env).
3) Provar que (sempre ocorre na FE, mas não necessariamente, o modelo precisa pertencer à FE):

$$a(\theta)[t(\boldsymbol{x}) - \tau(\theta)] = \frac{\partial}{\partial\theta} \ln f_{\boldsymbol{X}}(\boldsymbol{x};\theta)$$

$\mathcal{E}(t(\boldsymbol{X})) = \tau(\theta), \forall\theta \in \Theta.$

Obtenção de Estatísticas Suficientes e Completas

1) Família exponencial regular (FEC não, necessariamente, apresenta uma estatística suficiente e completa).
2) Critério da fatoração (ou teorema da minimalidade e suficiência) & provar a completitude pela respectiva definição.

Exemplo

- Seja $X_1, ..., X_n$ uma aa de X, $X \sim \exp(\theta)$, $\mathcal{E}(X) = \theta$. Defina $\tau(\theta) = e^{-1/\theta}$. Obtenha o ENVVUM de $\tau(\theta)$.
- Note que:

$$F_X(x;\theta) = 1 - e^{-x/\theta} \rightarrow P(X > 1) = 1 - F_X(1;\theta) = e^{-1/\theta}.$$

- Por outro lado, defina:

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

- Assim, $\mathcal{E}_\theta(T) = P(X_1 > 1) = e^{-1/\theta}$.
- Segundo o TLS (TRB), $T^* = \mathcal{E}(T|S)$ é o ENVVUM de $\tau(\theta)$, em que $S = \sum_{i=1}^n X_i$.

Exemplo

- Mas, note que:

$$\mathcal{E}(T|S=s) = P(X_1 > 1|S=s) = \int_1^\infty f(x_1|s)dx_1 = \int_1^\infty \frac{f(x_1,s)}{f(s)}dx_1 \quad (10)$$

- Vamos encontrar $f(x_1,s)$. Para isso, defina $Y_1 = X_1$, $Y_2 = X_1 + T_2 = S$, $T_2 = \sum_{i=2}^n X_i$. Por outro lado,

$$f_{\boldsymbol{Y}}(y_1,y_2) = |\boldsymbol{J}| f_{(X_1,T_2)}(y_1, y_2 - y_1)\mathbb{1}_A(\boldsymbol{y})$$

item Assim $x_1 = y_1$, $t_2 = y_2 - y_1$, $\rightarrow$

$$\boldsymbol{J} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$$

Exemplo

- Portanto,

$$
\begin{aligned}
f_{\boldsymbol{Y}}(y_1, y_2) &= f_{X_1}(y_1) f_{T_2}(y_2 - y_1) \mathbb{1}_{(0,\infty)}(y_1) \mathbb{1}_{(y_1,\infty)}(y_2) \\
&= \frac{1}{\theta} e^{-y_1/\theta} \frac{1}{\theta^{n-1}\Gamma(n-1)} e^{\frac{-(y_1+y_2)}{\theta}} (y_2 - y_1)^{n-2} \mathbb{1}_{(0,y_2)}(y_1) \\
&\times \mathbb{1}_{(0,\infty)}(y_s) \\
&= \frac{1}{\theta} e^{-y_2/\theta} \frac{1}{\theta^{n-1}\Gamma(n-1)} (y_2 - y_1)^{n-2} \mathbb{1}_{(0,y_2)}(y_1) \mathbb{1}_{(0,\infty)}(y_s)
\end{aligned}
$$

- Dessa forma,

$$
f_{Y_1|Y_2}(y_1|y_2) = \frac{n-1}{y_2^n} \left(\frac{y_2 - y_1}{y_2} \right)^{n-2} \mathbb{1}_{(0,y_2)}(y_1)
$$

$$
f_{X_1|S}(x_1|s) = \frac{n-1}{s^n} \left(\frac{s - x_1}{s} \right)^{n-2} \mathbb{1}_{(0,s)}(x_1)
$$

Exemplo

- Assim,

$$
\begin{aligned}
\mathcal{E}(T|S=s) &= \int_1^{\infty} \frac{n-1}{s^{n-1}}(s-x_1)^{n-2}\mathbb{1}_{(0,s)}(x_1)dx_1 \\
&= \frac{n-1}{s^{n-1}}\int_1^{s}(s-x_1)^{n-2}dx_1 \underbrace{=}_{u=s-x_1} \frac{(s-1)^{n-1}}{s^{n-1}} \\
&= \left(1-\frac{1}{s}\right)^{n-1}, s>1
\end{aligned}
$$

Exemplo

- Logo,

$$\mathcal{E}(T|S) = \begin{cases} \left(1 - \frac{1}{S}\right)^{n-1}, & \text{se } s = \sum_{i=1}^{n} x_i > 1 \\ 0, & \text{caso contrário} \end{cases}$$

- Exercício: Seja $X_1, ..., X_n$ uma aa de $X_i \sim$ Poisson(λ_i), $\lambda_i = e^{\theta z_i}$, $z_i; i = 1, 2, ..., n$ são conhecidos. Encontre o $LICR(\tau(\theta)) = \theta$.

Definição

- Def: Sejam $\widehat{\theta}_1$ e $\widehat{\theta}_2$ dois estimadores não viciados de $\tau(\theta)$. Suponha que suas variâncias sejam ambas finitas. A eficiência de $\widehat{\theta}_1$ relativa a $\widehat{\theta}_2$ é definida como sendo:

$$\text{eff}_\theta = \frac{\mathcal{V}_\theta(\widehat{\theta}_1)}{\mathcal{V}_\theta(\widehat{\theta}_2)}, \forall \theta \in \Theta$$

- Dizemos que $\widehat{\theta}_1$ é mais eficiente do que $\widehat{\theta}_2$ se eff ≤ 1, $\forall \theta \in \Theta \leftrightarrow \mathcal{V}_\theta(\widehat{\theta}_1) \leq \mathcal{V}_\theta(\widehat{\theta}_2)$, $\forall \theta \in \Theta$.

Definição

- Def: Seja $\widehat{\theta}$ um env de $\tau(\theta)$ e suponha que $f_{\boldsymbol{X}}(.;\theta), \theta \in \Theta$, satisfaz as condições de regularidade. Dizemos que $\widehat{\theta}$ é um estimador eficiente se

$$\mathcal{V}(\widehat{\theta}) = \frac{[\tau'(\theta)]^2}{I(\theta)}, \forall \theta \in \Theta$$

ou seja, atinge o $LICR(\tau(\theta))$.

Definição

- Def: Seja $\widehat{\theta}$ um estimador eficiente (se existir). A eficiência de um estimador não viciado, $\widehat{\theta}_1$ (se existir) é definido como:

$$\text{eff}_\theta(\widehat{\theta}_1) = \frac{\mathcal{V}_\theta(\widehat{\theta}_1)}{\mathcal{V}_\theta(\widehat{\theta})}, \forall \theta \in \Theta.$$

Caso não exista um estimador eficiente (como na definição acima) definimos a eficiência como:

$$\text{eff}_\theta(\widehat{\theta}_1) = \frac{\mathcal{V}_\theta(\widehat{\theta}_1)}{LICR(\tau(\theta))}, \forall \theta \in \Theta$$

desde que $f_{\boldsymbol{X}}(.;\theta)$ satisfaça as CR.

- OBS: Seja $t(\boldsymbol{X})$ um env de $\tau(\theta)$, $\mathcal{V}_\theta(t(\boldsymbol{X})) = LICR(\tau(\theta))$ e $f_{\boldsymbol{X}}(.;\theta)$ satisfazendo as CR. Então $\boldsymbol{X}$ pertence à FE e $t(\boldsymbol{X})$ é suficiente. Prova: Bickel & Docksum (2015).

Definição

- Obs: O TLS também é válido no caso multiparamétrico (multivariado), ou seja $\boldsymbol{\theta} \in \Theta \subseteq \mathcal{R}^k$. Seja $\tau(\boldsymbol{\theta}) : \Theta \rightarrow \mathcal{R}$, $\boldsymbol{S} = (S_1, ..., S_k)'$ uma estatística suficiente e completa. Seja $T = g(\boldsymbol{S})$, $\mathcal{E}(T) = \tau(\theta)$. Então T é o ENVVUM de $\tau(\theta)$.
- Exemplo: Seja $X_1, ..., X_n$ uma aa de $X \sim N(\mu, \sigma^2)$. Sabemos que $\boldsymbol{S} = (\overline{X}, \sum_{i=1}^{n}(X_i - \overline{X})^2)'$ é uma estatística suficiente e completa.
- Então, seja $\widehat{\mu} = \overline{X}$ e $\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.
- Como $\mathcal{E}(\widehat{\mu}) = \mu$ e $\mathcal{E}(\widehat{\sigma}^2) = \sigma^2$, e ambos são funções de $\boldsymbol{S}$. Logo $\widehat{\mu}$ e $\widehat{\sigma}^2$ são os ENVVUM's de μ e σ^2, respectivamente, pelo TLS.