INFERÊNCIA ESTATÍSTICA (EST0035)

AULA03: ESTIMAÇÃO PONTUAL

Frederico Machado Almeida
frederico.almeida@unb.br

Departamento de Estatística
Instituto de Exatas
Universidade de Brasília (UnB)

Estimação Pontual

- Nas aulas anteriores foi apresentado o conceito de *estatística*, *estimador e estimativa*, bem como as respectivas propriedades que caracterizam um bom estimador.
- Porém, nada foi discutido sobre como tais estimadores e/ou estatísticas podem ser obtidos(as).
- Nessa parte da disciplina serão apresentados alguns dos principais métodos de obtenção de estimadores pontuais.

Observação 1

A lógica que está por trás da estimação pontual é bastante simples. Ou seja, quando a amostragem é feita a partir de uma população descrita por uma fdp (ou fp) $f(x|\theta)$, o conhecimento que se tem sobre $\theta \in \Theta$, gera o conhecimento da população inteira.

Estimação Pontual

Definição 1 (Estimação)

Estimação é o nome técnico para o processo que consiste em utilizar os dados amostrais para estimar os valores de parâmetros populacionais (que são desconhecidos).

- *Essencialmente qualquer característica de uma população pode ser estimada a partir de uma amostra aleatória;*

A ideia geral por trás da estimação pontual é a seguinte:

- Quando a amostragem é feita a partir de uma população descrita por uma função $f(x, \theta)$, o conhecimento de θ a partir da amostra, gera todo o conhecimento para a população.

Estimação Pontual

Definição 2 (Estimador Pontual)

Um estimador pontual $\hat{\theta}_n$ *de* θ *é qualquer função* $T(X_1, \cdots, X_n)$ *da aa, que não depende do parâmetro que está sendo estimado. Cujo seu suporte deve necessariamente coincidir com o suporte de* θ*.*

- A definição 2 nos remete a concluir que um estimador é uma estatística (ou deve ser função de uma estatística).
- É de suma importância que entendamos a diferença entre um estimador e uma estimativa. Lembrando que, enquanto o estimador é uma função de uma *aa*, digamos $X_1, \cdots, X_n$, uma estimativa seria o valor (constante), que o estimador pode assumir depois que a amostra for observada. Isto é, $X_1 = x_1, \cdots, X_n = x_n$.

Estimação Pontual

- Em alguns casos, é muito fácil obter o estimador de um parâmetro de interesse θ.Em outros casos (*modelos mais complexos*), nem tanto.
- Alguns dos principais métodos de estimação pontual incluem:
  1. Método dos Momentos (MM).
  2. Método de Máxima Verossimilhança (MMV).
  3. Métodos dos Mínimos Quadrados (MMQ).
  4. Métodos Bayesianos (MBayes).

Método dos Momentos

- O MM é provavelmente um dos procedimentos mais antigos para encontrar estimadores pontuais.
- Datando pelo menos nos finais do século XIX, este método é creditado ao estatístico inglês *Karl Pearson*.
- Em geral esse método tem a virtude de ser bastante simples em sua utilização, e quase sempre, gera algum tipo de estimador pontual.

Método dos Momentos

Definição 3 (Momentos Populacionais)

Seja $X \sim f(X|\theta)$ uma va e $c \in \mathbb{R}$ um escalar. Então, o r-ésimo momento de X centrado em c é dado por:

$$\mu_r^{'} = \mathbb{E}\left[(X-c)^r\right] = \int_{\mathcal{A}} (X-c)^r f(X|\theta)\, dX. \tag{1}$$

- Segue da equação 1 que, se $c = \mu$ obtemos o r-ésimo momento de X centrado na média.
- Assim, segue imediatamente que,

$$\begin{aligned} \mu_1^{'} &= \mathbb{E}\left[(X-c)\right] \\ \mu_2^{'} &= \mathbb{E}\left[(X-c)^2\right] \\ \cdots & \quad \cdots \\ \mu_d^{'} &= \mathbb{E}\left[(X-c)^d\right]. \end{aligned}$$

Método dos Momentos

- Se $c = 0$ obtemos os momentos centrados em zero, que em geral, são os momentos que estamos interessados em computar.
- Desta forma, segue da equação 1 que, os momentos centrados no ponto $c = 0$ são,

$$\mu_r = \mathbb{E}\left(X^r\right) = \int_{\mathcal{A}} X^r f\left(X|\theta\right) dX. \tag{2}$$

- Consequentemente, obtemos os seguintes momentos:

$$\begin{aligned}
\mu_1 &= \kappa_1\left(\theta\right) = \mathbb{E}\left(X\right) \\
\mu_2 &= \kappa_2\left(\theta\right) = \mathbb{E}\left(X^2\right) \\
\mu_3 &= \kappa_3\left(\theta\right) = \mathbb{E}\left(X^3\right) \\
\ldots & \quad \ldots \quad \ldots \\
\mu_d &= \kappa_d\left(\theta\right) = \mathbb{E}\left(X^d\right)
\end{aligned}$$

Método dos Momentos

Definição 4 (Momentos Amostrais)

Seja $X_1, X_2, \cdots, X_n$ uma aa extraída de uma população caracterizada por uma fdp $f(X|\theta)$. O r-ésimo momento amostral de X é dado por:

$$m_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r. \tag{3}$$

Observe que, o momento amostral de ordem $r = 1$ é dado pela média amostral. Isto é, $m_1 = \sum\limits_{i=1}^{n} X_i/n = \bar{X}_n$.

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i,\ m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2,\ m_3 = \frac{1}{n}\sum_{i=1}^{n} X_i^3,\ \cdots,\ m_d = \frac{1}{n}\sum_{i=1}^{n} X_i^d.$$

Método dos Momentos

Definição 5 (Estimador dos Momentos)

Seja $X_1, \cdots, X_n$ uma aa proveniente de uma população caraterizada por uma fdp (ou fp) $f(X|\theta)$. Um estimador dos momentos (EMom) $\hat{\theta}_n^*$ de θ, é obtido igualando os momentos amostrais, com os populacionais, isto é, $m_r = \mu_r$, com $r = 1, 2, \cdots, d$.

- Em linhas gerais, o EMom é obtido resolvendo o seguinte sistema de equações:

$$
\begin{aligned}
m_1 &= \mu_1 \Longrightarrow \frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}(X) \\
m_2 &= \mu_2 \Longrightarrow \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mathbb{E}\left(X^2\right) \\
\cdots &\quad \cdots \\
m_d &= \mu_d \Longrightarrow \frac{1}{n}\sum_{i=1}^{n} X_i^d = \mathbb{E}\left(X^d\right).
\end{aligned}
$$

Método dos Momentos

A aplicação do MM requer basicamente seguir os seguintes passos:

1. Identificar qual(is) parâmetro(s) deve(m) ser estimado(s).
2. Obter os primeiros r momentos populacionais μ_r.
3. Obter os primeiros r momentos amostrais m_r.
4. Igualar o r-ésimo momento amostral com o populacional, isto é, $m_r = \mu_r$, e de seguida resolver a equação resultante com o intuito de obter o EMom.
5. A solução advinda da equação do passo 4 corresponde o EMom para θ.

Exemplo 1: Assuma que $X_1, X_2, \cdots, X_n$ denota uma *aa* extraída de uma distribuição exponencial, isto é, $X \sim Exp(\theta)$, com *fdp* dada por: $f(x|\theta) = \theta e^{-\theta x}$, com $\theta > 0$ e $x \geq 0$. Pede-se para,

a) Obter o estimador dos momentos para θ.

b) Obter uma estimativa para o EMom, supondo a seguinte amostra observada: $x_1 = 3$, $x_2 = 7$ e $x_3 = 5$.

Método dos Momentos

Exemplo 2: Assuma que $X_1, X_2, \cdots, X_n$ denota uma *aa* proveniente de uma distribuição normal, tal que, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, com *fdp* dada por: $f\left(x|\mu, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left(x-\mu\right)^2\right\}$, com $\sigma^2 > 0$, $\mu \in \mathbb{R}$ e $x \in \mathbb{R}$. Obtenha os EMom para μ e σ^2.

Exemplo 3: Seja $X_1, X_2, \cdots, X_n$ uma *aa* de $X \sim \mathcal{U}\left(-\theta, \theta\right)$, com $\theta \in \mathbb{R}$. Obtenha o EMom $\hat{\theta}_n^*$ de θ, e encontre a $\mathbb{E}\left(\hat{\theta}_n^*\right)$.

Método dos Momentos

Exemplo 4: Seja $X_1, X_2, \cdots, X_n$ uma *aa* extraída de uma distribuição Gama (α, β), tal que, sua fdp tem a seguinte configuração:

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

a. Se $\mathbb{E}(X) = \frac{\alpha}{\beta}$ e $\mathbb{E}(X^2) = \frac{\alpha(\alpha+1)}{\beta^2}$, obtenha os EMom para α e β.
b. Considere a amostra observada: 152, 115, 109, 94, 88, 137, 152, 77, 160, 165, 125, 40, 128, 123, 136, 101, 62, 153, 83 e 69. Obtenha as estimativas para α e β, respectivamente.

Método de Máxima Verossimilhança

- O *método de máxima verossimilhança (MMV)*, é talvez um dos procedimentos (técnica) de estimação mais popular para derivar estimadores pontuais.
- Os estimadores obtidos por meio da aplicação do MMV são conhecidos como *estimadores de máxima verossimilhança (EMV)*.

Definição 6 (Função de Verossimilhança)

Seja $X_1, X_2, \cdots, X_n$ uma aa extraída de de uma população $\mathcal{F} = \{f(x|\theta); \theta \in \Theta\}$, ou família caracterizada por uma fdp (ou fp) $f(x|\theta)$. Então, a função de verossimilhança (ou densidade conjunta) é dada por:

$$f(x_1, \cdots, x_n|\theta) = L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\theta), \quad com \quad \theta \in \Theta. \tag{4}$$

Método de Máxima Verossimilhança

- A função de verossimilhança, ou densidade conjunta, é a mesma para todo $x_i \in \mathcal{X}$. Sendo $\mathcal{X}$ o espaço amostral de $X_1, \cdots, X_n$.
- Ou seja, a densidade muda apenas quando o valor do parâmetro de interesse, digamos θ for diferente. Portando, podemos simplificar a notação, denotando a função de verossimilhança como $L(\theta)$.

Definição 7

Seja $X_1, X_2, \cdots, X_n$ uma aa proveniente de uma população caracterizada por uma fdp (ou fp) $f(x|\theta)$. Assim, para cada ponto amostral $\mathbf{x}$, seja $\hat{\theta}_n = \hat{\theta}(\mathbf{x})$ um valor de θ que maximiza $L(\theta)$. Então, o valor $\hat{\theta}_n$ onde $L(\theta)$ atinge seu ponto de máximo é conhecido como EMV de θ.

Método de Máxima Verossimilhança

- Observe que, a suposição de independência entre as observações facilita a construção da função de verossimilhança dada em (4).
- Apesar disso, a maximização de $L(\theta)$ requer obter derivadas em cadeia de $f(x_i|\theta)$ para cada $i \in \{1, 2, \cdots, n\}$, o que não é uma tarefa fácil.
- Uma forma razoável de maximizar (4) consiste em utilizar alguma transformação que preserve a informação, e ao mesmo tempo, torne as derivadas mais simples de serem computadas.
- Assim, a transformação logarítmica permite transformar o produtório de (4) em soma. Ou seja,

$$\ell(\theta) = \log L(\theta) = \log\left[\prod_{i=1}^{n} f(x_i|\theta)\right] = \sum_{i=1}^{n} \log f(x_i|\theta). \tag{5}$$

Método de Máxima Verossimilhança

- Portanto, sem perda de generalidade, temos que, o valor de θ que maximiza a $L(\theta)$, será o mesmo que vai maximizar $\ell(\theta)$. Isto é,

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmax}}\, L(\theta) \Longrightarrow \hat{\theta}_n = \underset{\theta}{\operatorname{argmax}}\, \ell(\theta)$$

- Com base na log-verossimilhança apresentada na equação (5), o EMV será obtido considerando os seguintes passos:
  - *i* Obter a(s) função(ões) escore e igualá-las a zero. Ou seja, obter a(s) equação(ões) de verossimilhança.
  - *ii* Resolver a equação escore de (i) por forma a obter as raízes da equação. Tais raízes constituem os EMVs, de $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_d)^\top$.
  - *iii* Obter a(s) derivada(s) de segunda ordem, e verificar se $\ell''\left(\hat{\theta}_n\right) < 0$, para todo n. Se isso acontecer, significa que $\hat{\theta}_n$ denota um ponto de máximo global.

Método de Máxima Verossimilhança

- A equação escore é obtida igualando a zero, a primeira derivada de $\ell(\theta)$ com respeito a θ. Isto é,

$$\ell^{'}(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = 0, \text{ ou simplesmente } \mathcal{U}(\theta) = 0. \tag{6}$$

- Assim o EMV para θ será a solução da equação (6).
- É importante salientar que, para modelos mais complexos, pode não ser possível obter de forma analítica, a solução da equação (6). Nesses casos, métodos numéricos como devem ser considerados.
- Ademais, para aferimos se de fato $\hat{\theta}_n$ é um máximo global, precisamos avaliar se,

$$\left.\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}_n} < 0.$$

Método de Máxima Verossimilhança

Observação 2

1. *É importante enfatizar que, em situações onde o espaço paramétrico Θ é discreto, ou em que o máximo de $\ell(\theta)$ ocorre na fronteira de Θ, o EMV pode não ser obtido como solução da equação (6).*
2. *Nesses casos, o EMV de θ pode ser obtido simplesmente fazendo uma avaliação do comportamento da função de verossimilhança, com o intuito de entender onde tal função atinge o seu máximo.*

Exemplo 5: Sejam $X_1, X_2, \cdots, X_n$ variáveis aleatórias *iid*'s de uma distribuição Bernoulli(θ), com $\theta \in \Theta = \{\theta : 0 < \theta < 1\}$. Tal que, sua fmp é dada por:

$$\mathbb{P}(X = x|\theta) = \theta^x (1-\theta)^{1-x}, \text{ com } x = 0, 1.$$

Obtenha o EMV para o parâmetro de interesse θ.

Exemplo 6 (Motivação): Considere uma caixa que contem bolas brancas e vermelhas, tal que, sabe-se de antemão que a proporção θ de bolas vermelhas na caixa é de $1/3$ ou $2/3$. Ou seja, $\Theta = \{1/3, 2/3\}$.
Para obtermos informações sobre θ, uma *aa* de tamanho $n = 3$ foi observada com reposição, e apresentou bola vermelha na primeira extração, bola branca na segunda e na terceira extração. Se X_i denota uma *va*, tal que,

$$X_i = \begin{cases} 1 & \text{se a } i\text{-ésima bola foi vemelha} \\ 0 & \text{se a } i\text{-ésima bola foi branca.} \end{cases}$$

Assim, pede-se para obter o EMV $\hat{\theta}_n$ de θ.

Exemplo 7: Uma vez mais, considere o enunciado do Exemplo 2. Pede-se para obter o EMV para os parâmetros μ e σ^2, respectivamente.

Método de Máxima Verossimilhança

Exemplo 8: Seja $X_1, X_2, \cdots, X_n$ uma *aa* proveniente de uma distribuição uniforme, tal que, $X \sim \mathcal{U}(0, \theta)$, tal que, sua *fdp* é dada por:

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

Obtenha o EMV para θ.

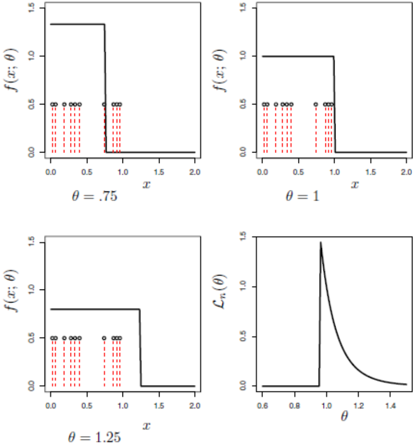

Figura: Função de verossimilhança de uma va $X \sim \mathcal{U}(0, \theta)$. Os primeiros 3 gráficos mostram a $f(x|\theta)$ para diferentes valores de θ. Observe que, quando $X_{(n)} > \theta$, a $f(x|\theta) = 0$, caso contrário, $f(x|\theta) = 1/\theta$.

Método de Máxima Verossimilhança

Propriedade dos EMV: como qualquer outro estimador pontual, os EMV satisfazem boa parte das propriedades desejáveis para um bom estimador, a saber: o EMV $\hat{\theta}_n$ de θ,

1. É equivariante. Isto é, se $\hat{\theta}_n$ é o EMV para θ, então, seja $g(\cdot)$ uma função qualquer, então, $\hat{\eta} = g\left(\hat{\theta}_n\right)$ denota o EMV para $g(\theta)$.
2. É em geral consistente. Isto é, $\hat{\theta}_n \xrightarrow{p} \theta_0$, com $\theta_0 \in \Theta$.
3. É assintoticamente normal, isto é, $\hat{\theta} \overset{a}{\sim} \mathcal{N}\left(\theta_0, Var\left(\hat{\theta}_n\right)\right)$.
4. É assintoticamente ótimo, ou eficiente. A grosso modo, isso significa que, entre todos os estimadores bem comportados, o EMV $\hat{\theta}_n$ de θ apresenta menor variância para grandes amostras.
5. Se aproxima da estimador de Bayes (ponto que será discutido no tópico de estimadores Bayesianos).

Método de Máxima Verossimilhança

- Em algumas situações de interesse (algumas apresentadas nas aulas anteriores), nosso interesse não se restringe em obter o EMV para θ, mas sim, de uma função função de θ, digamos $g(\theta)$.
- A obtenção do EMV para $g(\theta)$ é feita por meio de uma propriedade de suma importância em Estatística, conhecida como *propriedade de equivariância*.
- Informalmente falando, a *propriedade de equivariância* dos EMV's afirma que, se $\hat{\theta}_n$ é o EMV para θ, então, $g\left(\hat{\theta}_n\right)$ é um EMV para $g(\theta)$.

Método de Máxima Verossimilhança

Teorema 1 (Invariância dos EMV's)

Seja $X_1, X_2, \cdots, X_n$ uma aa proveniente de uma distribuição caracterizada pela fdp de X_i $f(X_i|\theta)$, com $\theta \in \Theta$. Para alguma função específica $g(\cdot)$, seja $\eta = g(\theta)$ uma função do parâmetro de interesse. Supondo que $\hat{\theta}_n$ denota o EMV para θ, então, $\hat{\eta} = g\left(\hat{\theta}_n\right)$ representa o EMV para $\eta = g(\theta)$.

Prova: Supondo $g(\cdot)$ uma função 1-a-1, segue que, $\eta = g(\theta) \Longrightarrow \theta = g^{-1}(\eta)$. Portanto, se $\hat{\theta}_n = \underset{\theta}{\operatorname{argmax}}\, L(\theta)$ é o EMV para θ, então, da equação (4) obtemos:

$$L(\eta|\mathbf{x}) = \prod_{i=1}^{n} f\left(x_i|g^{-1}(\eta)\right) = L\left(g^{-1}(\eta)\,|\mathbf{x}\right) = L\left(g^{-1}(\hat{\eta})\right).$$

Método de Máxima Verossimilhança

- Portanto, o EMV para $\eta = g(\theta)$ pode ser dado por:

$$\hat{\eta} = \underset{\eta = g(\theta)}{\operatorname{argmax}} \, L\left(g^{-1}(\eta)\right) = \underset{\eta = g(\theta)}{\operatorname{argmax}} \, L(\theta).$$

- Para finalizar a prova, observe que, como o máximo de $L(\theta)$ ocorrê no ponto $\hat{\theta}_n = g^{-1}(\hat{\eta})$, segue que, $\hat{\eta} = g\left(\hat{\theta}_n\right)$ é o EMV para $g(\theta)$.

Método de Máxima Verossimilhança

Exemplo 9: Seguindo o *Exemplo 8*, pode ser de interesse estimar a variância $Var(X) = \theta^2/12$. Então, sendo $X_i \sim \mathcal{U}(0, \theta)$, com $X_{(n)} = \max\{X_1, \cdots, X_n\}$ denotando o EMV, segue imediatamente que, $\hat{\eta} = g\left(\hat{\theta}_n\right) = g\left(X_{(n)}\right) = X_{(n)}^2/12$ é o EMV para $g(\theta) = \theta^2/12$.

Exemplo 10: Seja $X_1, \cdots, X_n$ uma *aa* de uma *va* $X \sim \mathcal{N}(\mu, 1)$. Vimos anteriormente que $\hat{\mu}_n = \bar{X}_n$ é o EMV de μ. Então, supondo que queremos estimar a função $g(\mu) = \mathbb{P}(X \leq 0) = \Phi(-\mu)$. Assim, segue do princípio de equivariância que, $g\left(\bar{X}_n\right) = \Phi\left(-\bar{X}_n\right)$ é o EMV para $g(\mu)$.

Teorema 2 (Consistência dos EMV's)

Assuma que $X_1, X_2, \cdots, X_n$ denota uma aa que satisfaz determinadas condições de regularidade, onde $\theta_0 \in \Theta$ denota o verdadeiro valor de θ, e $f(x|\theta)$ supõem-se ser uma função diferenciável com relação a θ. Então, a equação $\frac{\partial}{\partial\theta} L(\theta) = 0$ ou de forma equivalente, $\frac{\partial}{\partial\theta} \ell(\theta) = 0$ tem solução $\hat{\theta}_n$, que é tal que,

1. $\lim_{n \to \infty} \mathbb{E}_\theta (T_n) = \theta.$
2. $\lim_{n \to \infty} Var_\theta (T_n) = 0.$

Método de Máxima Verossimilhança

Exemplo 11: Considere novamente o enunciado do *Exemplo 5*, onde as va's $X_1, X_2, \cdots, X_n$ são *iid*'s e extraídas de uma distribuição Bernoulli(θ), agora com $\theta \in \Theta = \{\theta : 1/3 < \theta < 1\}$. Tal que, sua fp é dada por:

$$\mathbb{P}\left(X = x|\theta\right) = \theta^{x}\left(1-\theta\right)^{1-x}, \text{ com } x = 0, 1.$$

a. Para esse novo caso, obtenha o EMV para o parâmetro de interesse θ.
b. Assuma que a $X_1 = 1$, $X_2 = 0$, $X_3 = 0$, $X_4 = 0$ uma amostra observada. Obtenha a estimativa para $\hat{\theta}_n$.

Limite Inferior de Rao-Cramér

- Depois que os EMVs são obtidos, podemos de seguida computar sua variância (ou informação de Fisher) dos mesmos.
- Ademais, a avaliação da eficiência e eficiência assintotística dos EMV's requer a obtenção do limite inferior de Rao-Cramér (LIRC), por meio da desigualdade de informação.
- Para tal, antes de avançarmos com essa parte da disciplina, vamos estabelecer mais algumas condições de regularidade.

Limite Inferior de Rao-Cramér

Condições de Regularidade

Seja X uma va com *fdp* (ou *fp*) $f(x|\theta)$, com $\theta \in \Theta$, sendo Θ um intervalo aberto. As seguintes condições de regularidades podem ser acrescidas às definidas anteriormente:

- R0 As *fdp*'s (ou *fp*'s) são distintos, isto é, se $\theta_1 \neq \theta_2$ então, $f(x|\theta_1) \neq f(x|\theta_2)$.
- R1 As *fdp*'s (ou *fp*'s) tem o mesmo suporte para todo $\theta \in \Theta$.
- R2 O ponto θ_0 mora no interior do espaço paramétrico. isto é, $\theta_0 \in \Theta$.
- R3 A *fdp* (ou *fp*) $f(x|\theta)$ é duas vezes diferenciável com relação a θ.
- R4 A integral $\int f(x|\theta)\,dx$, é duas vezes diferenciável sob o sinal de integração com relação à θ.

Limite Inferior de Rao-Cramér

- Em linhas gerais, As condições de regularidades $(R0)$-$(R4)$ indicam que o suporte de X não pode depender de θ.
- Elas também indicam que podemos trocar a integral pela derivada com relação à θ.
- Sem perda de generalidade, assuma que as condições de regularidades apresentadas anteriormente sejam satisfeitas. Assim, a informação de Fisher de uma *va* X é dada por:

$$I(\theta) = \underbrace{\mathbb{E}_\theta\left[\left(\frac{\partial \ell(\theta)}{\partial \theta}\right)^2\right]}_{(i)} = \underbrace{\mathbb{E}_\theta\left[-\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right]}_{(ii)}. \tag{7}$$

Limite Inferior de Rao-Cramér

- É importante salientar que, em geral, a parte (ii) da equação (7) é a mais fácil de manusear, do que a (i). Sendo assim, a informação de Fisher pode ser resumida da seguinte forma:

$$I(\theta) = \mathbb{E}_\theta \left[-\frac{\partial^2 \ell(\theta)}{\partial \theta^2} \right]. \tag{8}$$

- A informação de Fisher pode ser interpretada como sendo uma média ponderada de $-\frac{\partial^2 \ell(\theta)}{\partial \theta^2}$, com relação aos pesos, representados por $f(x|\theta)$.

Exemplo 12: Seja X uma única variável com distribuição de Bernoulli(θ), tal que,

$$\mathbb{P}\left(X = x|\theta\right) = \begin{cases} \theta^{x}\left(1-\theta\right)^{1-x}, & \text{se } x = 0, 1 \\ 0, \text{ cc.} \end{cases}$$

Obtenha a informação de Fisher, $I_1\left(\theta\right)$.

Exemplo 13: Considere o caso em que a variável X segue uma distribuição normal com média μ e variância σ^2. Obtenha a matriz de informação de Fisher para a *va* X, digamos, $I_1\left(\mu, \sigma^2\right)$.

Limite Inferior de Rao-Cramér

- Observe que a informação de Fisher obtida nos Exemplos 12 e 13 foi considerada uma observação da *va* X (ou seja, amostra unitária).
- Portanto, em geral, vamos trabalhar com uma *aa* $X_1, X_2, \cdots, X_n$ extraída de uma população caracterizada por uma fdp (ou fp) $f(x|\theta)$. Nisso, surge a necessidade de generalizar o resultado apresentado anteriormente, para o caso em que temos uma aa com n va's.

Teorema 3 (Informação de Fisher baseada em uma *aa*)

Seja $X_1, X_2, \cdots, X_n$ uma aa tal que, para cada $i \in \{1, \cdots, n\}$ a informação de Fisher é dada por $I_1(\theta)$, então, a informação de Fisher referente a aa é dada por:

$$I_n(\theta) = nI_1(\theta).$$

Observe que, $I_n(\theta)$ denota a informação de Fisher baseada na aa completa, e $I_1(\theta)$ aquela baseada em X_1.

Limite Inferior de Rao-Cramér

Exemplo 14: Considere o resultado dos Exemplos 12 e 13. Obtenha a informação de Fisher, supondo que a *aa* $X_1, X_2, \cdots, X_n$ de X foi extraída de uma população, tal que *(i)* $X \sim Ber\,(\theta)$ e *(ii)* $X \sim \mathcal{N}\,(\mu, \sigma^2)$.
Mostre que, tanto faz usar o logaritmo da função densidade conjunta, ou apenas a densidade (ou fp) de X_1 (por meio do teorema 3), a informação de Fisher $I_n\,(\theta)$ é a mesma para os dois casos (*i*) e (*ii*).

Limite Inferior de Rao-Cramér

Teorema 4 (Desigualdade de Informação)

Seja $X_1, X_2, \cdots, X_n$ uma aa de uma va X é caracterizada por uma fdo/fp $f(x|\theta)$, com $\theta \in \Theta$. Assuma que as condições de regularidade $(R0)$-$(R4)$ sejam satisfeitas. E seja $T_n = T(X_1, \cdots, X_n)$ uma estatística quaisquer, tal que, $\mathbb{E}_\theta(T_n) = \mathbb{E}_\theta[T(X_1, \cdots, X_n)] = m(\theta)$. Então, o limite inferior de Rao-Cramér (LIRC) garante que,

$$Var(T_n) \geq \frac{[m'(\theta)]^2}{I_n(\theta)} = \frac{[m'(\theta)]^2}{nI_1(\theta)} \tag{9}$$

Nota: a prova será omitida aqui!!!

- A desigualdade de informação garante que, caso a variância do estimador T_n atinja o $LIRC$, então, $Var(T_n) = \frac{[m'(\theta)]^2}{I_n(\theta)}$.

Limite Inferior de Rao-Cramér

Corolário 1

Supondo que as condições de regularidade $(R0)$-$(R4)$ *são satisfeitas, e com base no resultado do teorema 4, segue que, se* T_n *é um estimador não-viesado de* θ*, isto é,* $m(\theta) = \theta$*, então, o LIRC é tal que,*

$$Var(T_n) \geq \frac{1}{I_n(\theta)},$$

pois, $m'(\theta) = 1$*.*

Limite Inferior de Rao-Cramér

Definição 8 (Estimador Eficiente)

Se T_n denota um estimador não-viesado para θ, então, a estatística T_n é dita ser um estimador eficiente de θ, se e somente se a variância de T_n atinge o LIRC, sob as condições usuais de regularidade.

- Em linhas gerais, um estimador T_n de θ é dito ser eficiente, se e somente se $LIRC\left(T_n\right)/Var_\theta\left(T_n\right) = 1$.
- É importante mencionar que, para alguns problemas, o EMV $\hat{\theta}_n$ pode ser não-viesado, porém, com variância superior ao LIRC. Nesses casos, o EMV é dito ser não eficiente.

Exemplo 15: uma vez mais, vamos considerar o resultado do Exemplo 5, onde uma *aa* $X_1, X_2, \cdots, X_n$ foi extraída de uma distribuição Bernoulli(θ). Vimos dos exemplos anteriores que, o EMV para θ é $T_n = \bar{X}_n$. Tal que, $\mathbb{E}_\theta(T_n) = \mathbb{E}_\theta\left(\bar{X}_n\right) = \theta$. De igual forma, sabe-se que $Var_\theta(T_n) = Var_\theta\left(\bar{X}_n\right) = \theta(1-\theta)/n$. Ademais, como $m(\theta) = \theta$, e $I_n(\theta) = \frac{n}{\theta(1-\theta)}$, segue imediatamente que, $LIRC = [m'(\theta)]^2 / I_n(\theta) = \frac{\theta(1-\theta)}{n}$, que é equivalente a variância de T_n. Então, concluímos que $Var(T_n)$ atinge o LIRC.

Limite Inferior de Rao-Cramér

Exemplo 16: Considere o caso em que a *aa* $X_1, X_2, \cdots, X_n$ foi extraída de uma distribuição uniforme, tal que, $X \sim \mathcal{U}(0, \theta)$, com fdp, $f(x|\theta) = 1/\theta$ para $0 < x < \theta$ e 0, caso contrário. Verifique se o EMV $X_{(n)}$ de θ é eficiente.

Exemplo 17: Seja $X_1, \cdots, X_n$ uma *aa* de $X \sim Poisson(\theta)$, com $\theta \in \Theta = \{\theta : \theta > 0\}$. Se $\hat{\theta}_n = \bar{X}_n$ denota o EMV para θ, mostre que $\hat{\theta}_n$ é um estimador eficiente.

Algoritmo Numéricos

Exemplo 18: Sejam $X_1, \cdots, X_n$ uma *aa* da distribuição da variável aleatória X com função de densidade dada por:

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

Nesse caso, a função de verossimilhança (ou densidade conjunta), é dada por:

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

Assim, o logaritmo da função de verossimilhança é dada por:

$$\ell(\theta) = -n\log(2) + \sum_{i=1}^{n}\log(1 + \theta x_i).$$

Consequentemente, $\mathcal{U}(\theta) = \ell'(\theta) = \sum_{i=1}^{n} \frac{x_i}{1+\theta x_i} = 0$, que não tem solução analítica.

Algoritmo Numéricos

- Conforme apresentado anteriormente, o método de estimação por máxima verossimilhança pressupõem encontrar o valor $\hat{\theta}_n$ que maximiza a função de verossimilhança, $\ell(\theta)$. Ou seja, a solução da equação,

$$\ell'(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = 0. \tag{10}$$

- Porém, há situações em que a equação (10) não pode ser resolvida analiticamente. Ou seja, quando não é possível obter estimadores em forma fechada.
- Nesses casos, métodos numéricos devem ser considerados para obter as raízes da equação escore.

Algoritmo Numéricos

- Diferentes métodos numéricos podem ser considerados para obter as estimativas de máxima verossimilhança para θ. Entre eles incluem, *o algoritmo de Newton-Raphoson (NR), o Escore de Fisher (FS), optim, etc.*
- A aplicação dos métodos de NR e FS requerem a obtenção da função escore, $\ell'(\theta)$ e a segunda derivada $\ell''(\theta)$.
- Os métodos baseados na maximização numérica requerem especificar um chute inicial para θ, digamos $\theta^{(0)}$, e a cada iteração, uma atualização de θ é obtida.

Algoritmo Numéricos

- O algoritmo de NR é um método iterativo onde a cada iteração, uma estimativa de θ é obtida. Em linhas gerais, tal método é baseado na seguinte equação:

$$\hat{\theta}^{(r+1)} = \hat{\theta}^{(r)} - \frac{\ell'\left(\hat{\theta}^{(r)}\right)}{\ell''\left(\hat{\theta}^{(r)}\right)}, \tag{11}$$

  com $r = 0, 1, 2, \cdots$.
- Observe que, para $r = 0$ temos o chute inicial. Feito isso, uma atualização na estimativa de θ, digamos, $\hat{\theta}^{(1)}$ é obtida. Isto é,

$$\hat{\theta}^{(1)} = \hat{\theta}^{(0)} - \frac{\ell'\left(\hat{\theta}^{(0)}\right)}{\ell''\left(\hat{\theta}^{(0)}\right)}.$$

- Na iteração 2 ($r = 1$) as quantidades $\ell'\left(\hat{\theta}^{(1)}\right)$ e $\ell''\left(\hat{\theta}^{(1)}\right)$ obtidas, e de seguida a equação (11).

Algoritmo Numéricos

- Assim, na iteração $r = 1$ obtemos a seguinte quantidade:

$$\hat{\theta}^{(2)} = \hat{\theta}^{(1)} - \frac{\ell'\left(\hat{\theta}^{(1)}\right)}{\ell''\left(\hat{\theta}^{(1)}\right)}.$$

- Este processo será repetido até obter a convergência do algoritmo de estimação. Ou seja, até que a quantidade $\left|\hat{\theta}^{(r+1)} - \hat{\theta}^{(r)}\right|$ seja menor que uma tolerância previamente definida. Ou seja, $\left|\hat{\theta}^{(r+1)} - \hat{\theta}^{(r)}\right| < \epsilon$.
- Outro critério de parada seria, parar o algoritmo de estimação se $\left|\ell\left(\hat{\theta}^{(r+1)}\right) - \ell\left(\hat{\theta}^{(r)}\right)\right| < \epsilon$.

Algoritmo Numéricos

- Observe que o algoritmo de estimação apresentado em (11) considera a informação observada, ou seja, usando $-\ell^{''}(\theta)$. Entretanto, o algoritmo FS usa a mesma ideia, com a diferença de que, ao invés de considerar a segunda derivada da função de verossimilhança, o FS considera a informação de Fisher. Isto é, $I_n(\theta) = \mathbb{E}(-\ell''(\theta))$.

$$\hat{\theta}^{(r+1)} = \hat{\theta}^{(r)} + \ell'\left(\hat{\theta}^{(r)}\right) I_n^{-1}\left(\hat{\theta}^{(r)}\right). \tag{12}$$

- Os critérios de convergência apresentados anteriormente, continuam valendo para o algoritmo de FS.

Método Delta

- Observe que os problemas discutidos nas aulas anteriores estão relacionados com situações onde a distribuição de uma variável aleatória.
- Em algumas situações, tem sido prática comum trabalhar (ou fazer inferência) para uma função de tais variáveis (ou estimadores).
- Tais transformações (ou funções dos estimadores) podem ser importantes para computar algumas quantidades de interesse, como: viés, variância, erro quadrático médio, etc.
- O método considerado para tal propósito é conhecido como método delta.

Método Delta

Definição 9

Suponha que $g(x)$ *é uma função diferenciável no ponto* $x \in \mathcal{X}$, *que denota o suporte da va* X. *Assim, se* T_n *denota um estimador de* $\theta \in \Theta$, *a expansão em série de Taylor em torno de* θ, *garante que:*

$$g(T_n) = g(\theta) + g'(\theta)(T_n - \theta).$$

Observe que, $\mathbb{E}[g(T_n)] = g(\theta)$, e $Var[g(T_n)] = [g'(\theta)]^2 \, Var(T_n)$.

Método Delta

Teorema 5

$\{T_n\}_{n\geq 1}$ *uma sequência de variáveis aleatórias, tal que,* $\sqrt{n}\left(T_n - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{T_n}\right)$, *com* $\sigma^2_{T_n} > 0$. *Assim, suponha que* $g\left(x\right)$ *é uma função contínua e diferenciável em* θ, *e que* $g^{'}\left(\theta\right) \neq 0$. *Então,*

$$\sqrt{n}\left(g\left(T_n\right) - g\left(\theta\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \left[g'\left(\theta\right)\right]^2 Var\left(T_n\right)\right).$$

Método Delta

Exemplo 19: Seja $X_1, \cdots, X_n$ uma aa proveniente de uma proveniente de uma distribuição Bernoulli(p), tal que, $\mu = p$ e $\sigma^2 = p\,(1-p)$. Seja $\hat{X}_n = \sum_{i=1}^{n} X_i/n$. Segue do Teorema 4, que $\bar{X}_n \overset{a}{\sim} \mathcal{N}\,(p, p(1-p)/n)$. Assim, suponha que seja possível uma função $g\,(\cdot)$ uma função diferenciável, tal que, $g\left(\bar{X}_n\right) = 1/\bar{X}_n$. Portando, usando o método delta é possível mostrar que,

$$g\left(\bar{X}_n\right) \overset{a}{\sim} \mathcal{N}\left(g\,(p)\,, \left[g'\,(p)\right]^2 Var\left(\bar{X}_n\right)\right),$$

onde $\bar{X}_n \equiv \hat{p}_n$, e $g\,(p) = 1/p$, respectivamente. Ou seja,

$$\sqrt{n}\left(\frac{1}{\bar{X}_n} - \frac{1}{p}\right) \overset{a}{\sim} \mathcal{N}\left(0, \frac{1-p}{p^3}\right).$$