Inferência Estatística II

Prof. Fernando de Souza Bastos
fernando.bastos@ufv.br

Departamento de Estatística
Programa de Pós-Graduação em Estatística Aplicada e Biometria
Universidade Federal de Viçosa
Campus UFV - Viçosa

Sumário

Limite Inferior de Cramér - Rao

O limite inferior de Cramér - Rao é o menor valor que a variância de um estimador pode assumir. Sob certas condições de regularidade mostraremos que os estimadores de Máxima Verossimilhança (EMV) atingem assintoticamente este limite inferior.

Mais Condições de Regularidade

Considere X uma v.a.c. com densidade $f(x,\theta)$, $\theta \in \Omega$, Ω é um intervalo aberto. Adicionalmente, considere também:

(R3) A densidade $f(x,\theta)$ é duas vezes diferenciável com relação a θ.

Mais Condições de Regularidade

Considere X uma v.a.c. com densidade $f(x, \theta)$, $\theta \in \Omega$, Ω é um intervalo aberto. Adicionalmente, considere também:

(R3) A densidade $f(x, \theta)$ é duas vezes diferenciável com relação a θ.

(R4) A integral $\int f(x, \theta)dx$ pode ser derivada duas vezes sob o sinal de integração como função de θ.

Obs: O caso discreto segue de forma similar.

Informação de Fisher

Considere a identidade:

$$\int_{-\infty}^{\infty} f(x,\theta)dx = 1 \Rightarrow \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f(x,\theta)dx = 0$$

Informação de Fisher

Considere a identidade:

$$\int_{-\infty}^{\infty} f(x,\theta)dx = 1 \Rightarrow \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f(x,\theta)dx = 0$$

$$\xrightarrow{\text{Sob } (R4)} \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(x,\theta)dx = 0$$

Informação de Fisher

Considere a identidade:

$$\int_{-\infty}^{\infty} f(x,\theta)dx = 1 \Rightarrow \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f(x,\theta)dx = 0$$

$$\xRightarrow{\text{Sob } (R4)} \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(x,\theta)dx = 0$$

$$\iff \int_{-\infty}^{\infty} \left\{\frac{\partial}{\partial\theta}\log f(x,\theta)\right\} f(x,\theta)dx = 0$$

Considere a identidade:

$$\int_{-\infty}^{\infty} f(x,\theta)dx = 1 \Rightarrow \frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f(x,\theta)dx = 0$$

$$\xRightarrow{\text{Sob } (R4)} \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(x,\theta)dx = 0$$

$$\iff \int_{-\infty}^{\infty} \left\{\frac{\partial}{\partial\theta}\log f(x,\theta)\right\} f(x,\theta)dx = 0$$

$$\iff E\left(\left\{\frac{\partial}{\partial\theta}\log f(x,\theta)\right\}\right) = 0 \qquad (\star)$$

Tomando novamente a derivada com relação a θ, obtemos:

$$\int_{-\infty}^{\infty} \left\{ \frac{\partial^2}{\partial\theta^2} \log f(x,\theta) \right\} f(x,\theta) dx + \int_{-\infty}^{\infty} \left\{ \frac{\partial}{\partial\theta} \log f(x,\theta) \right\} \frac{\partial}{\partial\theta} f(x,\theta) dx = 0 \qquad (\star\star)$$

Tomando novamente a derivada com relação a θ, obtemos:

$$\int_{-\infty}^{\infty}\left\{\frac{\partial^2}{\partial\theta^2}\log f(x,\theta)\right\}f(x,\theta)dx+\int_{-\infty}^{\infty}\left\{\frac{\partial}{\partial\theta}\log f(x,\theta)\right\}\frac{\partial}{\partial\theta}f(x,\theta)dx=0 \qquad (\star\star)$$

$$\int_{-\infty}^{\infty}\left\{\frac{\partial^2}{\partial\theta^2}\log f(x,\theta)\right\}f(x,\theta)dx+\int_{-\infty}^{\infty}\left\{\frac{\partial}{\partial\theta}\log f(x,\theta)\right\}\left\{\frac{\partial}{\partial\theta}\log f(x,\theta)\right\}f(x,\theta)dx$$

Tomando novamente a derivada com relação a θ, obtemos:

$$\int_{-\infty}^{\infty} \left\{\frac{\partial^2}{\partial\theta^2} \log f(x,\theta)\right\} f(x,\theta)dx + \int_{-\infty}^{\infty} \left\{\frac{\partial}{\partial\theta} \log f(x,\theta)\right\} \frac{\partial}{\partial\theta} f(x,\theta)dx = 0 \qquad (\star\star)$$

$$\int_{-\infty}^{\infty} \left\{\frac{\partial^2}{\partial\theta^2} \log f(x,\theta)\right\} f(x,\theta)dx + \int_{-\infty}^{\infty} \left\{\frac{\partial}{\partial\theta} \log f(x,\theta)\right\} \left\{\frac{\partial}{\partial\theta} \log f(x,\theta)\right\} f(x,\theta)dx$$

$$\int_{-\infty}^{\infty} \left\{\frac{\partial^2}{\partial\theta^2} \log f(x,\theta)\right\} f(x,\theta)dx + \int_{-\infty}^{\infty} \left\{\frac{\partial}{\partial\theta} \log f(x,\theta)\right\}^2 f(x,\theta)dx = 0$$

O segundo termo, denotado por $I(\theta)$, dado por

$$I(\theta) = \int_{-\infty}^{\infty} \left\{ \frac{\partial}{\partial\theta} \log f(x,\theta) \right\}^2 f(x,\theta)dx = E\left[\left(\frac{\partial}{\partial\theta} \log f(x,\theta) \right)^2 \right]$$

é conhecido como Informação de Fisher. Usando $(\star)$, temos que

$$I(\theta) = Var\left(\frac{\partial}{\partial\theta} \log f(x,\theta) \right).$$

Função Escore

Usando ($\star\star$), temos que

$$I(\theta) = -\int_{-\infty}^{\infty} \left\{ \frac{\partial^2}{\partial \theta^2} \log f(x,\theta) \right\} f(x,\theta) dx$$

Função Escore

Usando $(\star\star)$, temos que

$$I(\theta) = -\int_{-\infty}^{\infty} \left\{ \frac{\partial^2}{\partial \theta^2} \log f(x,\theta) \right\} f(x,\theta) dx$$

- $\dfrac{\partial}{\partial \theta} \log f(x,\theta)$ é conhecido como função Escore.

Função Escore

Usando $(\star\star)$, temos que

$$I(\theta) = -\int_{-\infty}^{\infty} \left\{ \frac{\partial^2}{\partial \theta^2} \log f(x, \theta) \right\} f(x, \theta) dx$$

- $\dfrac{\partial}{\partial \theta} \log f(x, \theta)$ é conhecido como função Escore.
- Um estimador de máxima verossimilhança é solução da equação

$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(x, \theta) = 0$$

Exemplo

Calcule $I(\theta)$

Suponha $X \sim \mathit{Bernoulli}(\theta)$ com $P(X = 0) = 1 - P(X = 1) = 1 - \theta$.

Exemplo

Calcule $I(\theta)$

Suponha $X \sim Bernoulli(\theta)$ com $P(X=0) = 1 - P(X=1) = 1-\theta$.

$$\log f(x,\theta) = x \log \theta + (1-x)\log(1-\theta)$$

$$\Rightarrow \frac{\partial^2}{\partial \theta^2} \log f(x,\theta) = -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}$$

$$\Rightarrow I(\theta) = -E\left(\frac{\partial^2}{\partial \theta^2} \log f(x,\theta)\right) = \frac{1}{\theta(1-\theta)}$$

Sejam $X_1, \dots, X_n$ v.a's iid com densidade $f(x, \theta)$ e $\boldsymbol{X} = (X_1, \dots, X_n)^T$. Nesse caso,

$$\frac{\partial \ell(\theta, \boldsymbol{X})}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(X_i, \theta)$$

Como cada termo na soma tem a mesma variância $I(\theta)$, segue que

$$\begin{aligned} Var\left(\frac{\partial \ell(\theta, \boldsymbol{X})}{\partial \theta}\right) &\overset{\text{iid}}{=} \sum_{i=1}^{n} \underbrace{Var\left(\frac{\partial}{\partial \theta} \log f(X_i, \theta)\right)}_{I(\theta)} \\ &= \sum_{i=1}^{n} I(\theta) = nI(\theta) \end{aligned}$$

Sejam $X_1, \ldots, X_n$ v.a's iid com densidade $f(x, \theta)$ e $\boldsymbol{X} = (X_1, \ldots, X_n)^T$. Nesse caso,

$$\frac{\partial \ell(\theta, \boldsymbol{X})}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(X_i, \theta)$$

Como cada termo na soma tem a mesma variância $I(\theta)$, segue que

$$Var\left(\frac{\partial \ell(\theta, \boldsymbol{X})}{\partial \theta}\right) \overset{\text{iid}}{=} \sum_{i=1}^{n} \underbrace{Var\left(\frac{\partial}{\partial \theta} \log f(X_i, \theta)\right)}_{I(\theta)}$$

$$= \sum_{i=1}^{n} I(\theta) = nI(\theta)$$

Assim, a informação em uma amostra aleatória de tamanho n é n vezes a informação em uma amostra de tamanho 1.

Limite Inferior de Cramér - Rao

Sejam $X_1, \ldots, X_n$ v.a's iid com densidade $f(x, \theta)$, $\theta \in \Omega$. Assuma que as condições de regularidade (R0) à (R4) valem. Seja $Y = u(X_1, \ldots, X_n)$ uma estatística com média $E(Y) = k(\theta)$. Então,

$$Var(Y) \geq \frac{[k^{'}(\theta)]^2}{nI(\theta)}$$

Demonstração

$$k(\theta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, \ldots, x_n) f(x_1; \theta) \cdots f(x_n; \theta) dx_1 \cdots dx_n$$

Diferenciando com respeito a θ, temos

$$k'(\theta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\partial u(x_1, \ldots, x_n)}{\partial \theta} f(x_1; \theta) \cdots f(x_n; \theta) dx_1 \cdots dx_n + \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, \ldots, x_n) \frac{\partial}{\partial \theta} \left[f(x_1; \theta) \cdots f(x_n; \theta) \right] dx_1 \cdots dx_n$$

Notem que,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\partial u(x_1, \ldots, x_n)}{\partial \theta} f(x_1; \theta) \cdots f(x_n; \theta) dx_1 \cdots dx_n = 0,$$

pois $u(x_1, \ldots, x_n)$ não depende de θ.

Logo,

$$k^{'}(\theta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, \ldots, x_n) \frac{\partial}{\partial \theta} \left[f(x_1; \theta) \cdots f(x_n; \theta) \right] dx_1 \cdots dx_n$$

$$\iff k^{'}(\theta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, \ldots, x_n) \frac{\partial}{\partial \theta} \prod_{i=1}^{n} f(x_i; \theta) dx_1 \cdots dx_n$$

$$\Rightarrow k^{'}(\theta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, \ldots, x_n) \sum_{i=1}^{n} \frac{\partial \log f(x_i, \theta)}{\partial \theta} \prod_{i=1}^{n} f(x_i; \theta) dx_1 \cdots dx$$

Seja $Z = \sum_{i=1}^{n} \frac{\partial \log f(x_i, \theta)}{\partial \theta}$. Logo,

$$k^{'}(\theta) = E(YZ) \qquad (\star\star\star)$$

Como visto anteriormente, $E(Z) = 0$ e $Var(Z) = nI(\theta)$, além disso, temos que a correlação entre Y e Z é dada por,

$$\rho = Cor(Y, Z) = \frac{COV(Y, Z)}{\sqrt{Var(Y)Var(Z)}}$$

$$COV(Y, Z) = E(YZ) - E(Y)E(Z)$$

Como $E(Z) = 0$, temos

$$\rho = \frac{E(YZ)}{\sigma_Y \sqrt{nI(\theta)}}$$

$$\Rightarrow E(YZ) = \rho \sigma_Y \sqrt{nI(\theta)} \qquad (\star\star\star\star)$$

Como visto anteriormente, $E(Z) = 0$ e $Var(Z) = nI(\theta)$, além disso, temos que a correlação entre Y e Z é dada por,

$$\rho = Cor(Y, Z) = \frac{COV(Y, Z)}{\sqrt{Var(Y)Var(Z)}}$$
$$COV(Y, Z) = E(YZ) - E(Y)E(Z)$$

Como $E(Z) = 0$, temos

$$\rho = \frac{E(YZ)}{\sigma_Y \sqrt{nI(\theta)}}$$
$$\Rightarrow E(YZ) = \rho \sigma_Y \sqrt{nI(\theta)} \qquad (\star\star\star\star)$$

Substituindo, $(\star\star\star\star)$ em $(\star\star\star)$, temos que

$$k^{'}(\theta) = \rho \sigma_Y \sqrt{nI(\theta)} \Rightarrow \frac{k^{'}(\theta)}{\sigma_Y \sqrt{nI(\theta)}} = \rho$$

Usando o fato de que $\rho^2 \leq 1$, temos

$$\frac{[k^{'}(\theta)]^2}{\sigma_Y^2 nI(\theta)} = \rho^2 \leq 1 \Rightarrow \sigma_Y^2 \geq \frac{[k^{'}(\theta)]^2}{nI(\theta)}$$

Corolário

Corolário:

Sob as condições do teorema anterior, se Y é um estimador não viesado de θ, ou seja, $k(\theta) = \theta$, então $Var(Y) \geq \dfrac{1}{nI(\theta)}$.

Corolário

Corolário:

Sob as condições do teorema anterior, se Y é um estimador não viesado de θ, ou seja, $k(\theta) = \theta$, então $Var(Y) \geq \dfrac{1}{nI(\theta)}$.

Definição 1

Seja Y um estimador não viesado de θ. A estatística Y é um estimador eficiente de θ se, e somente se, sua variância é igual ao limite inferior de Cramér-Rao.

Corolário

Corolário:

Sob as condições do teorema anterior, se Y é um estimador não viesado de θ, ou seja, $k(\theta) = \theta$, então $Var(Y) \geq \dfrac{1}{nI(\theta)}$.

Definição 1

Seja Y um estimador não viesado de θ. A estatística Y é um estimador eficiente de θ se, e somente se, sua variância é igual ao limite inferior de Cramér-Rao.

Definição 2

A razão entre o limite inferior de Cramér-Rao e a variância de um estimador, digamos Y, é chamada de eficiência do estimador Y.

Exemplo 1

$X_1, \ldots, X_n \sim$ Poisson(θ).

$$P(X_1 = k) = \frac{e^{-\theta}\theta^k}{k!}, \; \theta > 0, k = 0, 1, \ldots$$

O estimador de Máxima Verossimilhança de θ é $\hat{\theta} = \bar{X}$, $E(\bar{X}) = \theta$, $Var(\bar{X}) = \frac{\theta}{n}$. Logo, $I(\theta) = \frac{1}{\theta}$ e, portanto, o limite inferior de Cramér-Rao é $\frac{1}{n\frac{1}{\theta}} = \frac{\theta}{n}$. Como, $var(\bar{X})$ é igual ao limite inferior de Cramér-Rao, então $\bar{X}$ é um estimador eficiente para θ.

Exemplo 2

Contagem de Acidentes de Trânsito

Suponha que você esteja analisando a ocorrência de acidentes de trânsito em uma determinada interseção ao longo de um período de tempo. Você deseja modelar a distribuição das ocorrências de acidentes usando um modelo Poisson. O parâmetro que você está interessado em estimar é a taxa média de acidentes λ nessa interseção.

Exemplo 2

Contagem de Acidentes de Trânsito

Suponha que você esteja analisando a ocorrência de acidentes de trânsito em uma determinada interseção ao longo de um período de tempo. Você deseja modelar a distribuição das ocorrências de acidentes usando um modelo Poisson. O parâmetro que você está interessado em estimar é a taxa média de acidentes λ nessa interseção.

Durante um mês, você registra o número de acidentes que ocorreram diariamente na interseção. Os dados coletados são: $2, 1, 3, 0, 2, 4, 1, 2, 3, 1, 2, 0, 1, 2, 2, 3, 0, 1, 1, 2.$

Exemplo 2

O modelo Poisson é uma distribuição que descreve o número de eventos raros que ocorrem em um intervalo de tempo ou espaço fixo, quando esses eventos ocorrem de forma independente e com uma taxa média constante. A probabilidade de observar exatamente k eventos em um intervalo é dada pela fórmula da distribuição Poisson:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Exemplo 2

O parâmetro λ representa a taxa média de acidentes por dia. Você deseja estimar esse parâmetro com base nos dados coletados.

Exemplo 2

O parâmetro λ representa a taxa média de acidentes por dia. Você deseja estimar esse parâmetro com base nos dados coletados.

A função de verossimilhança é a expressão que relaciona a probabilidade dos dados observados com o parâmetro λ. Para um conjunto de dados, a função de verossimilhança é o produto das probabilidades de observar cada valor específico, dada a distribuição Poisson. A função de verossimilhança para os dados coletados é:

$$L(\lambda) = \frac{e^{-20\lambda}\lambda^{\sum k}}{k_1! \cdot k_2! \cdot \ldots \cdot k_n!}$$

onde $\sum k$ representa a soma dos valores dos acidentes observados e $k_1, k_2, \ldots, k_n$ são esses valores individuais.

Exemplo 2

Estimador de λ

O estimador de máxima verossimilhança para λ em um modelo Poisson é a média dos valores observados. Neste caso, a soma dos acidentes observados é 33 e o número de dias é 20, então o estimador de λ é:

$$\hat{\lambda} = \frac{\sum k}{n} = \frac{33}{20} = 1.65$$

Exemplo 2

Estimador de λ

O estimador de máxima verossimilhança para λ em um modelo Poisson é a média dos valores observados. Neste caso, a soma dos acidentes observados é 33 e o número de dias é 20, então o estimador de λ é:

$$\hat{\lambda} = \frac{\sum k}{n} = \frac{33}{20} = 1.65$$

Informação de Fisher

A informação de Fisher para um modelo Poisson é dada pelo inverso da variância da distribuição Poisson, que é λ. Portanto, a informação de Fisher é igual a $1/\lambda$.

$$\text{Informação de Fisher } (I(\lambda)) = \frac{1}{\hat{\lambda}} = \frac{1}{1.65} \approx 0.61$$

Exemplo 2

No exemplo específico, a informação de Fisher, indica a precisão da estimativa com base na sensibilidade da distribuição Poisson à variação da taxa média.

Exemplo 2

No exemplo específico, a informação de Fisher, indica a precisão da estimativa com base na sensibilidade da distribuição Poisson à variação da taxa média.

O ideal é comparar a Informação de Fisher de duas ou mais amostras distintas ou quando se deseja avaliar a precisão das estimativas em diferentes cenários. Ao comparar duas amostras distintas, a informação de Fisher pode ajudar a determinar qual amostra fornece informações mais úteis sobre os parâmetros em questão. Maior informação de Fisher indica uma amostra mais informativa, que proporciona estimativas mais precisas dos parâmetros.

Exemplo 2

Em resumo, se a informação de Fisher for alta para um determinado parâmetro, isso indica que os dados têm uma forte capacidade de distinguir diferentes valores desse parâmetro. Por outro lado, uma informação de Fisher baixa sugere que os dados têm menos poder para discernir variações no parâmetro, resultando em estimativas menos precisas.

Exemplo 2

Cálculo e interpretação da Função Escore

A função escore é a derivada da função de log-verossimilhança em relação ao parâmetro de interesse. No caso do modelo Poisson, a função de log-verossimilhança para os dados coletados é:

$$\ln L(\lambda) = -20\lambda + \sum k \ln(\lambda) - \ln(k_1!) - \ln(k_2!) - \ldots - \ln(k_n!)$$

A derivada da função de log-verossimilhança em relação a λ é a função escore:

$$\text{Escore}(\lambda) = -20 + \frac{\sum k}{\lambda}$$

Exemplo 2

Cálculo e interpretação da Função Escore

No contexto deste exemplo, a função escore expressa como a função de log-verossimilhança muda à medida que ajustamos o parâmetro λ, que representa a taxa média de acidentes. O valor da função escore nos diz em que direção e magnitude a probabilidade de observar os dados muda conforme alteramos λ.

Exemplo 2

Cálculo e interpretação da Função Escore

No cálculo acima, a função escore é $-20 + \frac{\sum k}{\lambda}$. Isso significa que quando você aumenta λ, o valor do escore diminui, indicando que a probabilidade de observar os dados diminui à medida que a taxa média de acidentes aumenta. Da mesma forma, quando você diminui λ, o valor do escore aumenta, indicando que a probabilidade de observar os dados aumenta à medida que a taxa média de acidentes diminui.

Mais uma condição de Regularidade

(R5) A densidade $f(x,\theta)$ é três vezes diferenciável com relação a θ. Além disso, para todo $\theta \in \Omega$ suponha a existência de uma constante c e uma função $M(x)$ tais que:

$$\left|\frac{\partial^3 \log f(x,\theta)}{\partial\theta^3}\right| \leq M(x), \text{ em que } E_{\theta_0}(M(X)) < \infty$$

para todo $\theta \in (\theta_0 - c, \theta_0 + c)$ com x no suporte de X.

Teorema

Teorema 1

Assuma que $X_1, \ldots, X_n$ é uma amostra aleatória de uma variável aleatória X com densidade $f(x, \theta_0)$, para $\theta_0 \in \Omega$ e que as condições de regularidade (R0) à (R5) valem. Além disso, assuma que $0 < I(\theta) < +\infty$. Então, qualquer sequência de soluções consistentes $\hat{\theta}_n$ da equação de estimação de Máxima Verossimilhança satisfaz:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, \frac{1}{I(\theta_0)})$$

$$\textbf{\textit{Obs:}} Var(\sqrt{n}\hat{\theta}_n) \approx \frac{1}{I(\theta_0)} \Rightarrow Var(\hat{\theta}_n) \approx \frac{1}{nI(\hat{\theta}_n)}, \; E(\hat{\theta}_n) = 0.$$

para $\hat{\theta}_n \sim N(0, \frac{1}{I(\theta_0)})$

Demonstração

Expandindo $\ell'(\hat{\theta}_n)$ em torno de θ_0 obtemos que

$$\ell'(\hat{\theta}_n) = \ell'(\theta_0) + (\hat{\theta}_n - \theta_0)\ell''(\theta_0) + \frac{(\hat{\theta}_n - \theta_0)^2}{2}\ell'''(\theta_n^*),$$

em que θ_0^* pertence ao intervalo entre θ_0 e $\hat{\theta}_n$. Como $\ell'(\hat{\theta}_n) = 0$, temos

$$0 = \ell'(\theta_0) + (\hat{\theta}_n - \theta_0)\ell''(\theta_0) + \frac{(\hat{\theta}_n - \theta_0)^2}{2}\ell'''(\theta_n^*)$$

$$\Rightarrow -\ell'(\theta_0) = \sqrt{n}(\hat{\theta}_n - \theta_0)\left[\frac{1}{\sqrt{n}}\ell''(\theta_0) + \frac{(\hat{\theta}_n - \theta_0)}{2\sqrt{n}}\ell'''(\theta_n^*)\right]$$

$$\Rightarrow \sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{-\ell'(\theta_0)}{\frac{1}{\sqrt{n}}\ell''(\theta_0) + \frac{(\hat{\theta}_n - \theta_0)}{2\sqrt{n}}\ell'''(\theta_n^*)}$$

$$\Rightarrow \sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{\frac{-\ell'(\theta_0)}{\sqrt{n}}}{\frac{1}{n}\ell''(\theta_0) + \frac{(\hat{\theta}_n - \theta_0)}{2n}\ell'''(\theta_n^*)}$$

Notem que:

1. $$\frac{-\ell'(\theta_0)}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta_0}\log f(x,\theta_0) \xrightarrow[n\to\infty]{TCL} N(0, I(\theta_0))$$

2. $$\frac{-\ell''(\theta_0)}{n} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta_0^2}\log f(x,\theta_0) \underset{\underset{LFGN}{\uparrow}}{\xrightarrow[n\to\infty]{P}} -E\left(\frac{\partial^2}{\partial\theta_0^2}\log f(x_i,\theta_0)\right)$$

$$= I(\theta_0)$$

Notem que:

3.

$$(\hat{\theta}_n - \theta_0) \xrightarrow{P} 0 \text{ e } \frac{1}{n}\ell'''(\hat{\theta}_n^*) \text{ é limitado em probabilidade. Daí}$$

Notem que:

3. $$(\hat{\theta}_n - \theta_0) \xrightarrow{P} 0 \text{ e } \frac{1}{n}\ell'''(\hat{\theta}_n^*) \text{ é limitado em probabilidade. Daí}$$

$$(\hat{\theta}_n - \theta_0)\frac{1}{n}\ell'''(\hat{\theta}_n^*) \xrightarrow{P} 0$$

Notem que:

3. $$(\hat{\theta}_n - \theta_0) \xrightarrow{P} 0 \text{ e } \frac{1}{n}\ell'''(\hat{\theta}_n^*) \text{ é limitado em probabilidade. Daí}$$

$$(\hat{\theta}_n - \theta_0)\frac{1}{n}\ell'''(\hat{\theta}_n^*) \xrightarrow{P} 0$$

Logo, por Slutsky,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, \frac{1}{I(\theta_0)}) \qquad \text{cqd} \blacksquare$$

Definição 3

$X_i \overset{iid}{\sim} f(x,\theta)$. Suponha que $\hat{\theta}_{1n} = \hat{\theta}_{1n}(X_1, \ldots, X_n)$ é um estimador de θ_0 tal que

$$\sqrt{n}(\hat{\theta}_{1n} - \theta_0) \xrightarrow{D} N(0, \sigma^2_{\hat{\theta}_{1n}}).$$

Então,

1. *A eficiência relativa (ou eficiência assintótica) de $\hat{\theta}_{1n}$ é definida como* $e(\hat{\theta}_{1n}) = \dfrac{\frac{1}{I(\theta_0)}}{\sigma^2_{\hat{\theta}_{1n}}}$
2. *O estimador $\hat{\theta}_{1n}$ é dito ser assintoticamente eficiente se* $e(\hat{\theta}_{1n}) = 1$.

3. Se $\hat{\theta}_{2n}$ é um outro estimador satisfazendo

$$\sqrt{n}(\hat{\theta}_{2n} - \theta_0) \xrightarrow{D} N(0, \sigma^2_{\hat{\theta}_{2n}}),$$

então a eficiência relativa assintótica de $\hat{\theta}_{1n}$ e $\hat{\theta}_{2n}$ é definida como

$$e(\hat{\theta}_{1n}, \hat{\theta}_{2n}) = \frac{\sigma^2_{\hat{\theta}_{2n}}}{\sigma^2_{\hat{\theta}_{1n}}}.$$

Exercícios

Para 🏠

Exercícios 6.2: 1, 3 (item 1), 7 (letras a, b e c), 8, 9, 10, 11, 12, 16

Referências I

BOLFARINE, Heleno; SANDOVAL, Mônica Carneiro. **Introdução à inferência estatística**. [S.l.]: SBM, 2001. v. 2.

CASELLA, George; BERGER, Roger L. **Statistical inference**. [S.l.]: Cengage Learning, 2021.

HOGG, RV; MCKEAN, J; CRAIG, AT. **Introduction to Mathematical Statistics**. Eighth Edition. [S.l.]: Pearson, 2019.