Estatística Matemática

Estimação Pontual - Aula 2

Prof. Paulo Cerqueira Jr - cerqueirajr@ufpa.br
Faculdade de Estatística - FAEST
Programa de Pós-Graduação em Matemática e Estatística - PPGME
Instituto de Ciências Exatas e Naturais - ICEN

https://github.com/paulocerqueirajr

Propriedades dos estimadores

Introdução

- Nas últimas subseções apresentamos métodos para obter estimadores pontuais.
- Investigaremos agora algumas propriedades que ajudam a decidir se um estimador é melhor que outro.
- Assim como a qualidade dos estimadores.

Erro Quadrático Médio (EQM)

Erro Quadrático Médio (EQM)

- Uma medida da aproximação de um estimador $t(X_1, \ldots, X_n)$ em relação o parâmetro alvo $\tau(\theta)$ é o que chamamos de EQM do estimador.

Definição 1 (EQM) Seja $T = t(X_1, \ldots, X_n)$ um estimador para $\tau(\theta)$. Então, $E\left[(T - \tau(\theta))^2\right]$ é dito ser o EQM de $T = t(X_1, \ldots, X_n)$.

Observação:

O subscrito θ no símbolo da esperança E_θ indica a p.d.f. da família sendo considerada, a partir da qual a amostra é proveniente.

Erro Quadrático Médio (EQM)

$$E_\theta\left[(T-\tau(\theta))^2\right] = E_\theta\left[(t(X_1,\ldots,X_n)-\tau(\theta))^2\right]$$

$$= \int\cdots\int (t(x_1,\ldots,x_n)-\tau(\theta))^2 f(x_1|\theta)\cdots f(x_n|\theta)dx_1\cdots dx_n$$

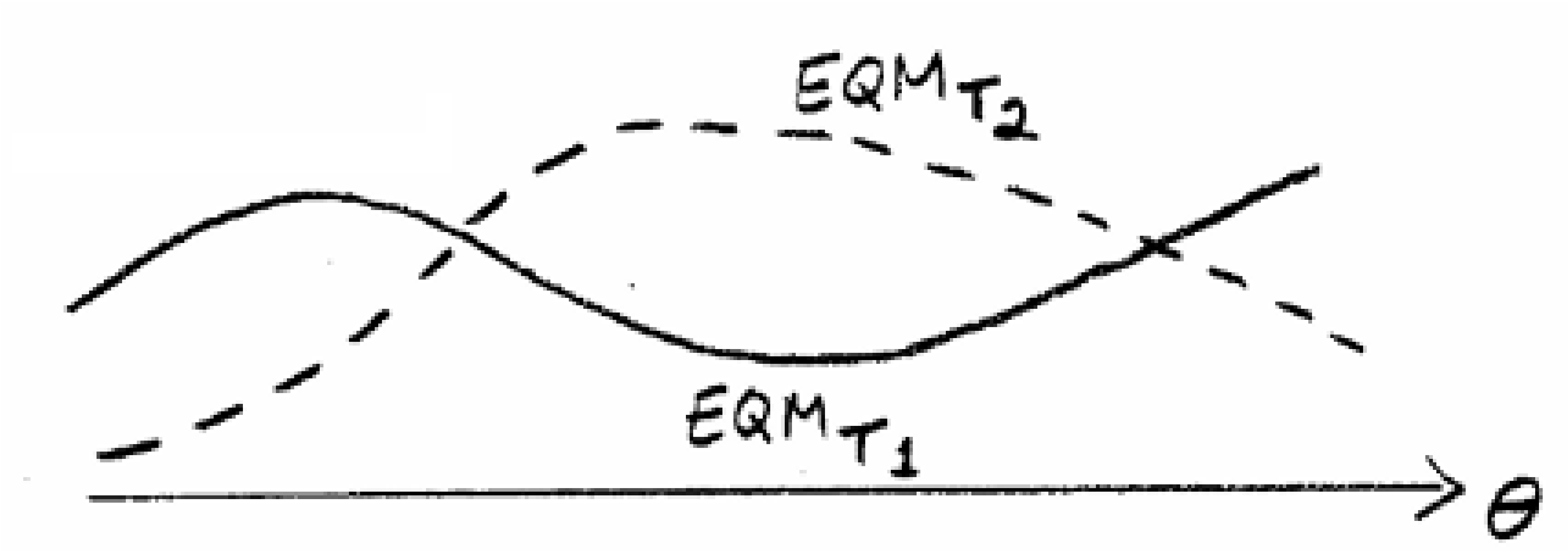
onde $f(x|\theta)$ é a p.d.f. da qual a amostra aleatória foi selecionada.

- O nome "Erro Quadrático Médio" pode ser justificado considerando o fato de que a diferença $t-\tau(\theta)$, sendo t o valor de T usado para estimar $\tau(\theta)$, representa o `erro na estimação` de $\tau(\theta)$.
- Trabalhamos com $(T-\tau(\theta))^2$ o que leva ao termo "Quadrático".
- O termo "Médio" está associado ao cálculo do valor esperado.
- Note que $E_\theta\left[(T-\tau(\theta))^2\right]$ é uma medida do distanciamento de T em relação a $\tau(\theta)$, assim como a variância de uma variável aleatória é a medida do distanciamento dos possíveis valores em relação a sua média.

Erro Quadrático Médio (EQM)

- Se formos comparar estimadores em termos do EQM, naturalmente iremos preferir aquele com menor EQM.
- Podemos pensar em obter um estimador com o menor EQM, mas tal estimador raramente existe.
- Em geral, o EQM de um estimador depende de θ.
- Para dois estimadores $T_1 = t_1(X_1, \ldots, X_n) \quad \text{e} \quad T_2 = t_2(X_1, \ldots, X_n)$ de $\tau(\theta)$, seus respectivos EQMs são funções de θ e podem se cruzar; ou seja, para algum θ temos:
  - T_1 com o menor EQM e para outros θ', temos
  - T_2 com o menor EQM.

Erro Quadrático Médio (EQM)



Erro Quadrático Médio (EQM)

- Em geral, qualquer função crescente da distância absoluta $|T-\tau(\theta)|$ serviria para medir a qualidade de um estimador.
- O "Erro Absoluto Médio" $E_\theta[|T-\tau(\theta)|]$ é uma alternativa razoável, mas o EQM tem pelo menos duas vantagens sobre outras medidas de distância:

  1. É mais fácil de manipular analiticamente.
  2. Possui a seguinte interpretação.

$$E_\theta[(W-\theta)^2] = \mathrm{Var}_\theta W + (E_\theta(W)-\theta)^2 = \mathrm{Var}_\theta W + [\mathrm{Viés}_\theta(W)]^2$$

- O resultado acima é obtido a partir de:

$$\mathrm{Var}_\theta(x) = \mathbb{E}(x^2) - \mathbb{E}^2(x)$$

`Note que:` $\mathrm{Var}_\theta(W-\theta) = \mathrm{Var}_\theta(W)$ pois θ é constante

$$\mathrm{Var}_\theta(W) = \mathrm{Var}_\theta(W-\theta) = \mathbb{E}_\theta[(W-\theta)^2] - \mathbb{E}_\theta^2[W-\theta]$$

Erro Quadrático Médio (EQM)

Definição 2 (Viés, Vício ou tendenciosidade.) Um estimador $T = t(X_{i_1}, \ldots, X_n)$ é dito não viciado para $T(\theta)$
se e somente se $E_\theta(T) = E_\theta[\bar{t}(X_{j_1}, \ldots, X_n)] = T(\theta)$ para todo $\theta \in \Theta$.

- O vício de um estimador pontual W de um parâmetro θ é dado pela diferença entre $E(W)$ e θ, isto é

$$\text{Vício}(W) = E_\theta(W) - \Theta$$

- Pode ser tanto positivo, negativo ou zero.
- Se o vício for igual a zero, dizemos que o estimador W é não viciado para θ.

Erro Quadrático Médio (EQM)

- Veja que:

$$EQM = E_\theta[(W-\theta)^2] = V(W) + [E_\Theta(W) - \theta]^2$$

- Incorpora dois componentes:
  - um medindo a variabilidade do estimador (precisão).
  - outro medindo seu vício.
- Um bom estimador terá valores pequenos para esses componentes.
- Para encontrar estimadores com boas propriedades de EQM, precisamos encontrar aqueles que controlam a variabilidade e o vício.
- Claramente, estimadores não viciados (vício = 0) serão preferidos.
- Para um estimador não viciado temos $E_\theta[(W-\theta)^2] = V(W)$, ou seja, o EQM é igual à variância do estimador.

Erro Quadrático Médio (EQM)

Exemplo 1 Exemplo: Seja $X_1, \ldots, X_n$ uma amostra aleatória da $N(\mu, \sigma^2)$. Sabemos que $\overline{X}$ e S^2 são estimadores não viciados, então $E(\overline{X}) = \mu \quad e \quad E(S^2) = \sigma^2$. Este resultado também é válido se não usarmos a suposição de normalidade dos dados.

EQMS:

- Considere $\theta = (\mu, \sigma^2) \quad \text{com ou sem a suposição de normalidade,}$

$$E_\theta[(X - \mu)^2] = Var(\bar{X}) = \sigma^2/n$$

$$E_\theta[(S^2 - \sigma^2)^2] = Var(S^2) = \frac{2\sigma^4}{\sqrt{n-1}}$$

- Lembre que:

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1}$$

Erro Quadrático Médio (EQM)

- Este resultado só é verdade se $X_1, \ldots, X_n$ são i.i.d. $N(\mu, \sigma^2)$.

$$Var(S^2) = Var\left(\frac{\sigma^2}{n-1}\frac{n-1}{\sigma^2}S^2\right) = \frac{\sigma^4}{(n-1)^2}Var\left(\frac{n-1}{\sigma^2}S^2\right) = 2(n-1)$$
$$= \frac{\sigma^4}{(n-1)^2}\cdot 2(n-1) = \frac{2\sigma^4}{\sqrt{n-1}}.$$

- A expressão do EQM para S^2 não permanece a mesma se retirarmos a suposição de normalidade.

Nota:

Embora muitos estimadores não viciados são também razoáveis em termos de EQM, esteja avisado que `controlar o vício não garante que o EQM seja controlado`.

Em particular, pode ocorrer uma troca entre a variância e o vício do estimador de tal forma que um aumento pequeno no vício leva a uma maior diminuição na variância, resultado em um EQM melhor (menor).

Erro Quadrático Médio (EQM)

Exemplo 2 $X_1, \ldots, X_n$ são i.i.d. $N(\mu, \sigma^2)$. Um estimador alternativo para σ^2 é o EMV dado por

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{x=1}^{n}(X_i - \overline{X})^2 = \frac{n-1}{n}S^2$$

- Veja que

$$E(\hat{\sigma}^2) = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}E(S^2) = \left(\frac{n-1}{n}\right)\sigma^2.$$

- Portanto, $\hat{\sigma}^2$ é um estimador viciado para σ^2.

Erro Quadrático Médio (EQM)

- A variância dos estimador:

$$V(\hat{\sigma}^2) = V\left(\frac{n-1}{n}S^2\right) = \left(\frac{n-1}{n}\right)^2 V(S^2) = \left(\frac{n-1}{n}\right)^2 \frac{2\sigma^4}{n-1}$$

$$= \frac{(n-1)\cdot 2\sigma^4}{n^2}$$

- Desta forma o EQM será: *[Math Processing Error]*.

[Math Processing Error]

Erro Quadrático Médio (EQM)

- Concluindo o exemplo...

[Math Processing Error]

- Ou seja, $\hat{\sigma}^2$ tem menor EQM que S^2. Se ignorarmos o problema do vício e usarmos $\hat{\sigma}^2$, iremos obter um EQM menor.
- O exemplo acima não implica que S^2 deva ser abandonado como estimador de σ^2.
- O argumento acima indica que, em média, $\hat{\sigma}^2$ será mais próximo de σ^2 do que S^2 se o EQM for usado como critério de comparação.
- Entretanto, $\hat{\sigma}^2$ é viciado e irá, em média, subestimar σ^2.
- Este fato é forte o bastante para nos deixar preocupados sobre o uso de $\hat{\sigma}^2$ como estimador de σ^2.

Consistência

Consistência no EQM

- As definições de EQM e Vício para um estimador consideram que o tamanho amostral n é fixo. A próxima propriedade a ser introduzida considera uma avaliação para a situação onde n cresce.
- Assuma que $T_n = t_n(X_1, \ldots, X_n)$ representa um estimador de $\tau(\theta)$ baseado em uma amostra de tamanho n. Iremos considerar uma sequência de estimadores:

$$T_1 = t_1(X_1)$$

$$T_2 = t_2(X_1, X_2)$$

$$T_3 = t_3(X_1, X_2, X_3)$$

$$\vdots$$

$$T_n = t_n(X_1, \ldots, X_n)$$

- Um exemplo óbvio é

$$T_n = t_n(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- As funções t_n na sequência serão o mesmo tipo de função para cada n.

- Quando consideramos uma sequência de estimadores, parece razoável pensar que uma boa sequência de estimadores deverá ter valores que se aproximam da quantidade a ser estimada conforme o tamanho amostral aumenta.

Consistência no EQM

Definição 3 (Consistência baseada no EQM.) Seja $T_1, T_2, \ldots, T_n, \ldots$ uma sequência de estimadores de $\tau(\theta)$, sendo $T_n = t_n(X_1, \ldots, X_n)$ baseado em uma amostra de tamanho n. Esta sequência de estimadores é dita "Consistente no EQM" para $\tau(\theta)$ se e somente se

$$\lim_{n\to\infty} E_\theta[(T_n - \tau(\theta))^2] = 0 \quad \text{para todo} \quad \theta \in \mathbb{R}$$

`Observação:` Consistência no EQM implica que ambos o vício e a variância de T_n se aproximam de zero visto que

$$E_\theta[(T_n - \tau(\theta))^2] = Var_\theta[T_n] + [\tau(\theta) - E_\theta(T_n)]^2$$

Consistência no EQM

Exemplo 3 Considere uma amostra aleatória obtida de uma p.d.f. com média μ e variância σ^2.

- Seja

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad e \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

- Considere as sequências $\overline{X}_1, \overline{X}_2, \overline{X}_3, \ldots \quad e \quad S_1^2, S_2^2, S_3^2, \ldots$
- Lembre que: $\overline{X}_n$ estima $\mu \quad e \quad S_n^2$ estima σ^2.
- Para $\overline{X}_n$:

$$EQM(\overline{X}_n) = E[(\overline{X}_n - \mu)^2] = Var(\overline{X}_n) = \frac{\sigma^2}{n}$$

$$\lim_{n\to\infty} E[(\overline{X}_n - \mu)^2] = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

Consistência no EQM

- Para S_n^2:

$$EQM(S_n^2) = E[(S_n^2 - \sigma^2)^2] = Var(S_n^2) = \frac{1}{n}\left[\mu_4 - \frac{n-3}{n-1}\sigma^4\right]$$

$$\lim_{n\to\infty}\left[\frac{\mu_4}{n} - \frac{n-3}{n-1}\sigma^4\right] = \lim_{n\to\infty}\frac{\mu_4}{n} - \lim_{n\to\infty}\frac{n-3}{n-1}\sigma^4 = 0 - 0 = 0$$

`Conclusão:` $\{\overline{X}_n\}$ e $\{S_n^2\}$ são sequências consistentes no EQM para seus respectivos parâmetros alvo.

- Note que se $T_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$, então, $\{T_n\}$ também será uma sequência consistente no EQM para σ^2.

Consistência no EQM

Definição 4 (Consistência simples.) Seja $T_1, T_2, T_3, \ldots, T_n, \ldots$ uma sequência de estimadores de $t(\theta)$, sendo $T_n = t_n(x_1, \ldots, x_n)$. A sequência $\{T_n\}$ é dita consistente (forma mais simples ou fraca) para $t(\theta)$ se para todo $\epsilon > 0$ temos

$$\lim_{n\to\infty} P[|T_n - t(\theta)| < \epsilon] = 1,$$

ou equivalentemente

$$\lim_{n\to\infty} P[|T_n - t(\theta)| \geq \epsilon] = 0.$$

- Note que a definição acima coincide com aquela que especificamos para a convergência em probabilidade. Esta definição tem relação com a lei fraca dos grandes números.
- Se um estimador é consistente no EQM, ele também atende a consistência simples. A relação contrária não necessariamente será verdade.

Consistência no EQM

- A prova considera a desigualdade de Chebychev estudada anteriormente.

$$P(|T_n - t(\theta)| \geq \epsilon) < E[(T_n - t(\theta))^2]/\epsilon^2.$$

- Se T_n é consistente no EQM, conforme $n \to \infty$,

$$E[(T_n - t(\theta))^2] \to 0 \Rightarrow P(|T_n - t(\theta)| \geq \epsilon) \to 0.$$

Melhores Estimadores Não Viciados

Melhores Estimadores Não Viciados

- Conforme visto anteriormente, uma comparação de estimadores baseada no EQM pode não determinar um favorito.
- Na verdade não existe um estimador com o “melhor EQM”.
- A razão disto é que a classe de todos os estimadores é muito grande.
- Por exemplo, o estimador $\hat{\theta} = 17$ não pode ser batido em termos de EQM para estimar $\theta = 17$, mas este é um estimador ruim se $\theta \neq 17$.
- Uma maneira de tornar mais acessível o problema de encontrar um “melhor estimador” é limitar a grande classe de estimadores.
- Uma forma de restringir a grande classe de estimadores será considerar apenas os estimadores não viciados.
- Se W_1 e W_2 são ambos estimadores não viciados do parâmetro θ, isto é, $E(W_1) = E(W_2) = \theta$, então seus EQMs serão iguais às suas variâncias.
- Portanto, iremos escolher o estimador com a menor variância. Este será o “melhor estimador não viciado”.
- Embora estejamos lidando com estimadores não viciados, os resultados apresentados aqui são na verdade mais gerais.

Melhores Estimadores Não Viciados

Definição 5 Um estimador W^* é dito o "Melhor Estimador Não Viciado" para $\tau(\theta)$ se satisfaz:

$$E_\theta(W^*) = \tau(\theta),$$

para todo θ e para qualquer outro estimador W com

$$E_\theta(W) = \tau(\theta),$$

temos

$$Var_\theta(W^*) \leq Var_\theta(W)$$

para todo θ.

- W^* é também chamado de `Estimador Não Viciado de Variância Uniformemente Mínima` (ENVVUM) para $\tau(\theta)$.
- Encontrar o `melhor estimador não viciado` (se existir) não é uma tarefa fácil por várias razões, duas delas são ilustradas a seguir.

Melhores Estimadores Não Viciados

Exemplo 4 Seja $X_1, \ldots, X_n$ i.i.d. Poisson(λ) e considere $\overline{X}$ e S^2 (a média e variância amostral). Lembre que na distribuição Poisson, tanto a média quanto a variância são iguais a λ.

- Então $E_\lambda(\overline{X}) = \lambda$ e $E_\lambda(S^2) = \lambda$ para todo λ.
- Logo: $\overline{X}$ e S^2 são estimadores não viciados para λ.
- Para determinar qual destes é o melhor estimador, iremos comparar suas variâncias.
- Sabemos que $Var_\lambda(\overline{X}) = \lambda/n.$.
- O cálculo de $Var_\lambda(S^2)$ é longo (iremos omitir aqui!).
- É possível mostrar que $Var_\lambda(\overline{X}) \leq Var_\lambda(S^2)$ para todo λ.
- Mesmo com o resultado acima estabelecendo que $\overline{X}$ é melhor que S^2, considere a seguinte classe de estimadores para λ:

$$W_\alpha(\overline{X}, S^2) = \alpha\overline{X} + (1-\alpha)S^2$$

Melhores Estimadores Não Viciados

- Mesmo que $\bar{X}$ seja melhor que S^2, seria ele $\bar{X}$ melhor que todo $W_\alpha(\bar{X}, S^2)$? Como podemos ter certeza de que não existe algum outro melhor estimador não viciado?
- Suponha que, para estimar o parâmetro $\tau(\theta)$ de uma distribuição $f(x|\theta)$, podemos especificar um limite inferior $\beta(\theta)$ para a variância de qualquer estimador não viciado para $\tau(\theta)$. Se podemos então encontrar um estimador não viciado W^* satisfazendo

$$Var_\theta(W^*) = \beta(\theta),$$

- teremos encontrado o melhor estimador não viciado.
- A abordagem acima considera o que chamamos de limite inferior de Cramér-Rao.

Melhores Estimadores Não Viciados

Teorema 1 (Desigualdade de Cramér-Rao.) Seja $X = (X_1, \ldots, X_n)$ uma amostra (variáveis não necessariamente independentes) com p.d.f. $f(x|\theta)$.

- Considere $W(X) = W(X_1, \ldots, X_n)$ como qualquer estimador satisfazendo

$$\frac{d}{d\theta} E_\theta[W(X)] = \int_X \frac{\partial}{\partial\theta} [W(x) f(x|\theta)]\, dx$$

e $Var_\theta[W(X)] < \infty$. Então,

$$Var_\theta[W(X)] \geq \frac{\left(\frac{d}{d\theta} E_\theta[W(X)]\right)^2}{E_\theta\left(\left[\frac{\partial}{\partial\theta} \log f(x|\theta)\right]^2\right)}$$

Melhores Estimadores Não Viciados

Corolário 1 (Desigualdade de Cramér-Rao (caso i.i.d.).) Se as suposições do Teorema 3.4.1 estão satisfeitas e além disso $X_1, \ldots, X_n$ são i.i.d. com p.d.f. $f(X|\theta)$, então

$$Var_\theta[W(X)] \geq \frac{\left(\frac{d}{d\theta}E_\theta[W(X)]\right)^2}{nE_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X_i|\theta)\right)^2\right]}$$

Melhores Estimadores Não Viciados

Observação 1: Embora o limite inferior de Cramér-Rao tenha sido apresentado para variáveis aleatórias contínuas, este resultado também é válido para o caso discreto. Se $f(x|\theta)$ é uma p.m.f., então devemos ser capazes de permutar a diferenciação e o somatório. Assumimos que mesmo $f(x|\theta)$ sendo uma p.m.f. não diferenciável em x, ela é diferenciável em θ (este é o caso das p.m.f.'s mais comuns).

Observação 2: A quantidade

$$E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right]$$

é chamada de **Informação de Fisher amostral**.

- Esta terminologia reflete o fato de que a Informação de Fisher fornece um limite para a variância do melhor estimador não viciado de θ. Conforme o valor da Informação de Fisher aumenta e obtemos mais informação sobre θ, teremos um limite menor para a variância do melhor estimador não viciado.

Notação:

Melhores Estimadores Não Viciados

Lema 1 Se $f(x|\theta)$ satisfaz

$$\frac{d}{d\theta} E_\theta \left[\frac{\partial}{\partial\theta} \log f(x|\theta) \right] = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial\theta} \log f(x|\theta) \right) f(x|\theta) dx$$

Isso será verdade para distribuições da família exponencial. Então

$$I_n(\theta) = E_\theta \left[\left(\frac{\partial}{\partial\theta} \log f(x|\theta) \right)^2 \right] = -E_\theta \left(\frac{\partial^2}{\partial\theta^2} \log f(x|\theta) \right)$$

Observação 3: Uma alternativa para o cálculo da Informação de Fisher é considerar o caso “observado” ao invés do “esperado”.

Melhores Estimadores Não Viciados

Informação de Fisher observada:

$$\hat{I}_n(\hat{\theta}) = -\frac{\partial^2}{\partial\theta^2} \log L(\theta; x)\Big|_{\theta=\hat{\theta}}$$

- Veja que aqui não aplicamos o cálculo do valor esperado.

Observação 4: Uma das suposições para aplicarmos o Teorema é a exigência de sermos capazes de permutar a derivada e a integral em $d/d\theta E_\theta[W(X)]$.

- As p.d.f.'s da família exponencial atendem esse critério. Se o domínio de $f(x|\theta)$ depende de θ, o Teorema não será apropriado (ex.: $f(x|\theta) = 1/\theta$ para $0 < x < \theta$).

Melhores Estimadores Não Viciados

Exemplo 5 $X_1, \ldots, X_n$ são i.i.d. Poisson(λ).

$$f(X_i|\lambda) = \frac{e^{-\lambda}\lambda^{X_i}}{x_i!} \in L(\lambda; X) = \prod_{i=1}^{n} f(X_i|\lambda)$$

- Temos

$$I_n(\lambda) = E\left[\left(\frac{\partial}{\partial\lambda}\log\prod_{i=1}^{n} f(X_i|\lambda)\right)^2\right] = -nE_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log f(X_i|\lambda)\right]$$

$$= -nE_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log\left(\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right)\right] = -nE_\lambda\left[\frac{\partial^2}{\partial\lambda^2}(-\lambda + x_i\log\lambda - \log x_i!)\right]$$

$$= -nE_\lambda\left[\frac{\partial}{\partial\lambda}(-1 + \frac{x_i}{\lambda})\right] = -nE_\lambda\left[-\frac{x_i}{\lambda^2}\right] = n\frac{1}{\lambda^2}E(X_i) = n\frac{\lambda}{\lambda^2} = \frac{n}{\lambda}$$

Melhores Estimadores Não Viciados

- Portanto, qualquer estimador W, não viciado para λ, terá

$$Var_{\lambda}(W) \geq \frac{1}{n/\lambda} = \frac{\lambda}{n}$$

Como W é não viciado,

$$\frac{d}{d\lambda} E_{\lambda}[W] = \frac{d}{d\lambda} \lambda = 1$$

- Lembre que $\overline{X}$ tem variância $\frac{\lambda}{n}$, então sua variância atinge o limite inferior de Cramér-Rao.
- $\overline{X}$ é o melhor estimador não viciado para λ.

Melhores Estimadores Não Viciados

Observação 5: Na busca do melhor estimador não viciado devemos ter em mente que, mesmo que o cálculo do limite inferior de Cramér-Rao seja aplicável, não existe garantia de que este limite seja igual à variância de algum estimador não viciado para aquele parâmetro de interesse. O limite inferior pode não ser atingível.

Exemplo 6 Sejam $X_1, \ldots, X_n$ i.i.d $N(\mu, \sigma^2)$ e considere o interesse em estimar σ^2, sendo μ desconhecido.

- A p.d.f. da Normal pertence à família exponencial. Veja que:

$$\frac{\partial^2}{\partial(\sigma^2)^2}\log\left[(2\pi\sigma^2)^{1/2}\exp\left\{-\frac{1}{2\sigma^2}(X_i-\mu)^2\right\}\right] = \frac{\partial^2}{\partial(\sigma^2)^2}\left[-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(X_i-\mu)^2\right]$$

$$= \frac{\partial}{\partial(\sigma^2)}\left[-\frac{1}{2}\frac{1}{\sigma^2} + \frac{(X_i-\mu)^2}{2\sigma^4}\right] = \frac{1}{2\sigma^4} - \frac{(X_i-\mu)^2}{\sigma^6}$$

Melhores Estimadores Não Viciados

$$-nE\left[\frac{\partial^2}{\partial(\sigma^2)^2}\log f(X_i,\mu,\sigma^2)\right] = -nE\left[\frac{1}{2\sigma^4} - \frac{(X_i-\mu)^2}{\sigma^6}\right]$$

$$= -n\left[\frac{1}{2\sigma^4} - \frac{1}{\sigma^6}E[(X_i-\mu)^2]\right]$$

$$= -n\left[\frac{1}{2\sigma^4} - \frac{\sigma^2}{\sigma^6}\right] = -n\left[\frac{1}{2\sigma^4} - \frac{1}{\sigma^4}\right] = -n\left[-\frac{1}{2\sigma^4}\right] = \frac{n}{2\sigma^4}.$$

- Portanto, qualquer estimador não viciado W para σ^2 deve satisfazer

$$\mathrm{Var}(W) \geq \frac{1}{n/2\sigma^4} = \frac{2\sigma^4}{n}.$$

Melhores Estimadores Não Viciados

- Sabemos que $\mathrm{Var}(S^2) = \frac{2\sigma^4}{n-1}$. Logo, S^2 não atinge o limite inferior de Cramér-Rao.
- Esta resposta nos deixa com a seguinte dúvida:

> Existe um estimador não viciado para σ^2 melhor do que S^2? Ou o limite inferior de Cramér-Rao não pode ser atingido neste caso?

- É possível mostrar que se μ é desconhecido, o limite inferior calculado no exemplo acima não pode ser atingido; ver Casella e Berger (2002), pag. 341.

Melhores Estimadores Não Viciados

- Até o momento, o conceito de suficiência não foi usado em nossa busca por estimadores não viciados.
- Veremos agora que a suficiência é na verdade uma ferramenta importante.

Teorema 2 (Teorema de Rao-Blackwell.) Seja W qualquer estimador não viciado de $\tau(\theta)$, e seja T uma estatística suficiente para θ. Defina $\phi(T) = E(W|T)$. Então $E[\phi(T)] = \tau(\theta)$ e $\mathrm{Var}_\theta[\phi(T)] \leq \mathrm{Var}_\theta[W]$ para todo θ, isto é, $\phi(T)$ é um Estimador Não-viciado Uniformemente Melhor para $\tau(\theta)$.

Melhores Estimadores Não Viciados

Exemplo 7 (Condicionando em uma estatística não suficiente.) Seja $X_1, X_2 \sim N(\theta, 1)$. A estatística $\overline{X} = (X_1 + X_2)/2$ apresenta

$$E(\overline{X}) = \theta \quad \text{e} \quad Var(\overline{X}) = \frac{1}{4}[Var(X_1) + Var(X_2)] = \frac{1}{4}(1+1) = \frac{1}{2}$$

- Condicionando em X_1 (que não é suficiente), seja $\phi(X_1) = E[\overline{X}|X_1]$.
- Temos então:

$$E[\phi(X_1)] = E[E(\overline{X}|X_1)] = E(\overline{X}) = \theta$$

$$Var[\phi(X_1)] = Var[E(\overline{X}|X_1)] = Var(\overline{X}) - E[Var(\overline{X}|X_1)] \leq Var(\overline{X})$$

- Logo $\phi(X_1)$ é melhor do que $\overline{X}$.

Melhores Estimadores Não Viciados

- Entretanto:

$$\phi(X_1) = E[\overline{X}|X_1] = \frac{1}{2}E[X_1|X_1] + \frac{1}{2}E[X_2|X_1]$$

$$= \frac{1}{2}X_1 + \frac{1}{2}\theta = \frac{1}{2}(X_1 + \theta)$$

- Visto que $E(X_2|X_1) = E(X_2) = \theta$ por independência.
- A formulação de $\phi(X_1)$ depende de θ, então $\phi(X_1)$ não é um estimador.

Melhores Estimadores Não Viciados

- Sabemos agora que, na busca pelo melhor estimador não viciado de $t(\theta)$, precisamos considerar apenas estimadores baseados em estatísticas suficientes.
- A questão será: se tivermos

$$E[\phi] = t(\theta)$$

e ϕ sendo baseado em uma estatística suficiente, isto é,

$$E(\phi|T) = \phi$$

como saberemos que ϕ é o melhor estimador não viciado?

- Claro que, se ϕ atinge o limite inferior de Cramér-Rao, então ele é o melhor estimador, entretanto, se ele não atinge o limite de Cramér-Rao teremos ganhado alguma coisa?

Teorema 3 Se W é um "melhor estimador não viciado" para $t(\theta)$, então W é único.

Teorema 4 Seja T uma estatística suficiente completa para θ, e seja $\varphi(T)$ qualquer estimador baseado apenas em T. Então $\varphi(T)$ é o único "melhor estimador não viciado" para seu valor esperado $E[\varphi(T)]$.

- O teorema de *Lehmann-Scheffé* representa um avanço importante em estatística-matemática unindo os conceitos de suficiência, completude e unicidade.
- Este teorema é uma junção dos dois últimos teoremas.

Teorema 5 (Teorema de *Lehmann-Scheffé*.) Estimadores não viciados baseados em estatísticas suficientes completas são únicos.

Eficiência

Eficiência

- A propriedade da `Consistência` está relacionada com a aproximação assintótica de um estimador.
- O estimador converge para o valor real do parâmetro alvo?
- Agora iremos estudar a propriedade de `Eficiência` que está focada na variância assintótica de um estimador.
- No cálculo da variância assintótica, considere a seguinte definição:
- Para um estimador, se

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \sigma^2)$$

sendo k_n uma sequência de constantes, então σ^2 é chamado de `variância limite` ou `limite das variâncias`.

Eficiência

Exemplo 8 Para a média $\overline{X}_n$ de n observações i.i.d. da Normal com $E(X) = \mu$ e $Var(X) = \sigma^2$.

Se $T_n = \overline{X}_n$, então

$$\lim_{n \to \infty} nVar(\overline{X}_n) = \lim_{n \to \infty} n\frac{\sigma^2}{n} = \sigma^2$$

é a variância limite de T_n.

Eficiência

Definição 6 Para um estimador T_n, suponha que

$$\sqrt{n}(T_n - \tau(\theta)) \xrightarrow{d} N(0, \sigma^2)$$

- O parâmetro σ^2 é chamado de variância assintótica ou variância da distribuição limite de T_n.
- É interessante notar que a variância assintótica é sempre menor que a variância limite. Não mostraremos este resultado, mas ele pode ser encontrado em Casella e Berger (2002), pag. 471.
- Considerando o limite inferior de Cramér-Rao, podemos determinar uma variância assintótica ótima.

Definição 7 Uma sequência de estimadores W_n é assintoticamente eficiente para um parâmetro $\tau(\theta)$ se

$$\nu(\theta) = \frac{[\tau'(\theta)]^2}{E\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right]}.$$

isto é, a variância assintótica de W_n atinge o limite inferior de Cramér-Rao (caso: 1 observação amostral).

Propriedades assintóticas dos EMVs

Propriedades assintóticas dos EMVs

1. Os EMVs são estimadores consistentes de seus parâmetros alvo.

- Para ter essa propriedade, a p.m.f. ou p.d.f., definida no problema, deverá satisfazer algumas "condições de regularidade".
  a. Observamos $X_1, \ldots, X_n$ sendo $X_i \sim f(x|\theta)$ i.i.d.'s;
  b. O parâmetro θ é identificável, isto é, se $\theta \neq \theta'$ então $f(x|\theta) \neq f(x|\theta')$;
  c. A p.m.f.s ou p.d.f.s $f(x|\theta)$ possuem domínio (suporte) comum, e $f(x|\theta)$ é derivável em relação a θ;
  d. O espaço paramétrico Θ contém um "conjunto aberto" A para o qual o verdadeiro valor θ_0 do parâmetro θ é um "ponto interior".
- Estas condições de regularidade estarão atendidas em todos os problemas estudados neste curso.
- Não precisaremos nos preocupar com elas aqui, mas é importante saber que elas existem.

Propriedades assintóticas dos EMVs

Teorema 6 (Consistência dos EMVs.) Sejam $X_1, X_2, X_3, \ldots$ i.i.d.'s com p.m.f. ou p.d.f. $f(x|\theta)$. Considere $L(\theta; X) = \prod_{i=1}^{n} f(x_i; \theta)$ a função de verossimilhança e $\hat{\theta}$ representa o EMV de θ.

- Defina $U(\theta)$ como uma função contínua de θ.
- Se as condições de regularidade **a**, **b**, **c** e **d** forem atendidas, então para todo $\xi > 0$ e todo $\theta \in \Theta$,

$$\lim_{n \to \infty} P[|U(\hat{\theta}) - U(\theta)| \geq \xi] = 0$$

isto é, $U(\hat{\theta})$ é um estimador consistente para $U(\theta)$.

Propriedades assintóticas dos EMVs

2. Os EMVs são estimadores assintoticamente eficientes.

- Para ter essa propriedade, a p.m.f. ou p.d.f., definida no problema, deverá satisfazer as condições a, b, c, d e também:

    e. Para todo $x \in X$, a p.m.f. ou p.d.f. $f(x|\theta)$ é três vezes derivável com respeito a θ. A terceira derivada é contínua em θ, e

$$\int f(x|\theta)\,dx$$

    pode ser derivada três vezes sob o sinal de integração.

Propriedades assintóticas dos EMVs

f. Para qualquer $\theta_0 \in \Theta$, existe um número positivo c e uma função contínua $M(x)$, tal que

$$\left| \frac{\partial^3}{\partial \theta^3} \log f(x|\theta) \right| \leq M(x) \text{ para todo } x \in X,$$

$$\theta_0 - c < \theta < \theta_0 + c \quad \text{com} \quad E_\theta[M(x)] < \infty$$

- Novamente, ressaltamos que estas condições de regularidade estarão atendidas para todos os problemas propostos neste curso.

Propriedades assintóticas dos EMVs

Teorema 7 (Eficiência assintótica dos EMVs.) Sejam $X_1, X_2, X_3, \ldots$ i.i.d.'s com p.m.f. ou p.d.f. $f(x|\theta)$. Denote $\hat{\theta}$ como o EMV de θ, e considere $U(\theta)$ uma função contínua de θ. Se as condições de regularidade **a**, **b**, **c**, **d**, **e** e **f** para $f(x|\theta)$ forem atendidas, então:

$$\sqrt{n}\left[U(\hat{\theta}) - U(\theta)\right] \xrightarrow{d} N(0, r(\theta))$$

sendo $r(\theta) = \frac{[U'(\theta)]^2}{I_1(\theta)}$ o limite inferior de Cramér-Rao (para 1 observação). Isto é, $U(\hat{\theta})$ é estimador consistente e assintoticamente eficiente para $U(\theta)$.

Propriedades assintóticas dos EMVs

- Equivalentemente, podemos escrever também:

$$\tau(\hat{\theta}) \sim N\left(\tau(\theta), \frac{[\tau'(\theta)]^2}{nI_1(\theta)}\right)$$

sendo $I_1(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X_i|\theta)\right)^2\right]$ a informação de Fisher (caso: 1 observação). Este resultado é válido visto que $X_1, \cdots, X_n$ são i.i.d.

> Conclusão: Em geral, podemos dizer que os EMVs são assintoticamente:

- **Eficientes:** sua variância atinge o limite inferior de Cramér-Rao (amostra de tamanho n);
- **Consistentes:** seu valor esperado se aproxima do valor real do parâmetro alvo (assintoticamente não viciado);
- **Normais:** a distribuição do estimador se aproxima da Normal.

Método Delta

Método Delta

- O Método Delta é uma generalização útil do TCL.
- Aplicável em situações onde estamos interessados na distribuição de uma função de uma variável aleatória.

Teorema 8 (Método Delta.) Seja $Y_n = Y_1, Y_2, Y_3, \ldots$ uma sequência de variáveis aleatórias satisfazendo $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$. Para uma dada função g e um valor específico de θ, suponha que $g'(\theta)$ exista e não é igual a zero,

$$\sqrt{n}[g(Y_n) - g(\theta)] \xrightarrow{d} N(0, \sigma^2[g'(\theta)]^2).$$

Método Delta.

Exemplo 9 Suponha que X_i é uma variável aleatória com $E(X_i) = \mu \neq 0$. Considere o estimador

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}.$$

- Quando n é grande, temos pelo TCL o resultado:

$$\overline{X}_n \sim N\left[\mu, \frac{Var(X_i)}{n}\right]$$

- O que podemos dizer da distribuição assintótica de $\frac{1}{\overline{X}_n}$, ou seja, $g(\overline{X}_n) = \frac{1}{\overline{X}_n}$.

> Usaremos o Método Delta:

$$g(\mu) = \frac{1}{\mu}, \quad g'(\mu) = -\frac{1}{\mu^2}, \quad Var(\overline{X}_n) = \frac{Var(X_i)}{n} \quad \text{e} \quad E(\overline{X}_n) = \mu$$

Método Delta.

Pelo TCL:

$$\sqrt{n}(\overline{X}_n - \mu) \longrightarrow N\left[0, Var(X_i)\right]$$

Pelo M. Delta:

$$\sqrt{n}\left[\frac{1}{\overline{X}_n} - \frac{1}{\mu}\right] \longrightarrow N\left[0, Var(X_i)\left(-\frac{1}{\mu^2}\right)^2\right]$$

Conclusão: Quando n é grande, a variável aleatória $\frac{1}{\overline{X}_n}$ terá aproximadamente distribuição

$$N\left[\frac{1}{\mu}, \frac{Var(X_i)}{n\mu^4}\right]$$

- Se desconhecemos μ e $Var(X_i)$, iremos utilizar o resultado acima com a substituição desses elementos pelos seus respectivos estimadores. Sendo assim, a variância aproximada seria $\frac{S^2}{n\overline{X}^4}$.

Método Delta.

Exemplo 10 Suponha que $X_1, \ldots, X_n$ são i.i.d. $f(x|\theta)$. Considere $\hat{\theta} = EMV$ de θ

$$I_n(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\log L(\theta; X)\right)^2\right]$$

- Informação de Fisher baseada na amostra $X = (X_1, \ldots, X_n)$.
- Desejamos avaliar resultados assintóticos para $g(\hat{\theta})$, sendo $g(\cdot)$ uma função tal que $g(\theta)$ existe e não é zero.
- Visto que $\hat{\theta}$ é EMV, $\sqrt{n}[\hat{\theta} - \theta] \xrightarrow{d} N[0, \nu(\theta)]$, sendo $\nu(\theta)$ o limite inferior de Cramér-Rao (1 obs.)

$$\nu(\Theta) = \frac{\left[\frac{d}{d\theta}E(\hat{\theta})\right]^2}{E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X_i|\theta)\right)^2\right]} = \frac{1}{I_1(\theta)} \Rightarrow n \text{ grande } \Rightarrow E(\hat{\theta}) \approx \theta$$

Método Delta.

- Variância assintótica de $\hat{\theta}$ é $\nu(\theta)$.
- Através do Método Delta temos

$$\sqrt{n}[g(\hat{\theta}) - g(\theta)] \longrightarrow N[0, \nu(\theta)[g'(\theta)]^2].$$

- Então a variância assintótica de $g(\hat{\theta})$ será

$$\nu(\theta) \cdot [g'(\theta)]^2 = [g'(\theta)]^2 / I_1(\theta).$$

- Consequentemente, a variância de $g(\hat{\theta})$ será $[g'(\theta)]^2 / I_n(\theta)$

Método Delta.

- Este último resultado pode ser aproximado usando a Informação de Fisher Observada.

$$I_n(\hat{\theta}) = -\frac{\partial^2}{\partial\theta^2}\log L(\theta; X)\Big|_{\theta=\hat{\theta}}$$

- Aproximação desejada: Variância de

$$g(\hat{\theta}) = \frac{[g'(\theta)]^2}{I_n(\theta)} \approx \frac{[g'(\theta)]^2\Big|_{\theta=\hat{\theta}}}{-\frac{\partial^2}{\partial\theta^2}\log L(\theta; X)\Big|_{\theta=\hat{\theta}}}$$

- Note que o processo de estimação da variância envolve 2 passos:
  1. obter a variância de $g(\hat{\theta})$,
  2. estimar essa variância, em geral, substituindo θ por $\hat{\theta}$.