MAE0311 - Inferência Estatística

Fernando Henrique Ferraz Pereira da Rosa

25 de agosto de 2003

Lista 1[1]

1.7 Mostre que o $\hat{N}_5$ no Exemplo 1.3.9 é um estimador não viciado para N.

$$\hat{N}_5 = \frac{X_{(n)}^{n+1} - (X_{(n)} - 1)^{n+1}}{X_{(n)} - (X_{(n)} - 1)^n}$$

Tomemos $g(x) = \frac{x^{n+1} - (x-1)^{n+1}}{x - (x-1)^n}$ e calculemos $E(g(X_{(n)})$:

$$\begin{aligned}
E(g(X_{(n)}) &= \sum_{k=1}^{N} g(x_{(n)}) P(X_{(n)} = k) \\
&= \sum_{k=1}^{N} \frac{k^{n+1} - (k-1)^{n+1}}{k^n - (k-1)^n} \left(\left(\frac{k}{N}\right)^n - \left(\frac{k-1}{N}\right)^n \right) \\
&= \sum_{k=1}^{N} \frac{k^{n+1} - (k-1)^{n+1}}{k^n - (k-1)^n} N^{-n} (k^n - (k-1)^n) \\
&= N^{n+1} N^{-n} \\
&= N
\end{aligned}$$

Logo $\hat{N}_5$ é não viciado para N.

1.9 Seja X uma única variável aleatória com distribuição de Bernoulli com parâmetro θ. Sejam $\hat{\theta}_1 = X$ e $\hat{\theta}_2 = 1/2$ dois estimadores de θ.

(a) Verifique se $\hat{\theta}_1$ e $\hat{\theta}_2$ são não viciados para θ.
Temos que se $X \sim Bernoulli(\theta)$ então $E(X) = \theta$ e $Var(X) = \theta(1 - \theta)$. Basta calcularmos as esperanças dos dois estimadores:

$$E(\hat{\theta}_1) = E(X) = \theta$$

$$E(\hat{\theta}_2) = E(1/2) = 1/2$$

Logo $\hat{\theta}_1$ é um estimador não viciado para θ, enquanto $\hat{\theta}_2$ é um estimador viciado para θ.

[1] Powered by $\LaTeX 2_\varepsilon$, R 1.7.1 and Gentoo 1.4

(b) Compare os EQMs. Faça um gráfico dos EQMs como função de θ.

Calculando os EQMs:

$$EQM(\hat{\theta}_1) = Var(\hat{\theta}_1) + B^2(\hat{\theta}_1) = Var(X) + 0 = \theta(1-\theta)$$

$$\begin{aligned} EQM(\hat{\theta}_2) &= Var(\hat{\theta}_2) + B^2(\hat{\theta}_2) = Var(1/2) + (E(1/2)-\theta)^2 \\ &= \theta^2 - \theta + 1/4 \end{aligned}$$

Na Figura 1 temos o gráfico em função de θ dos dois EQMs. Igualando os dois EQMs temos que os pontos de intersecção (indicados no gráfico) c_1 e c_2 são, respectivamente: $\left(\frac{2-\sqrt{2}}{4}, \frac{1}{8}\right)$ e $\left(\frac{2+\sqrt{4}}{4}, \frac{1}{8}\right)$. Analisando o gráfico, percebemos que:

$$\theta \in \left[0, \frac{2-\sqrt{2}}{4}\right] \cup \left[\frac{2+\sqrt{2}}{4}, 1\right] \Rightarrow \hat{\theta}_1 \text{é melhor}$$

Caso contrário, $\hat{\theta}_2$ é melhor.

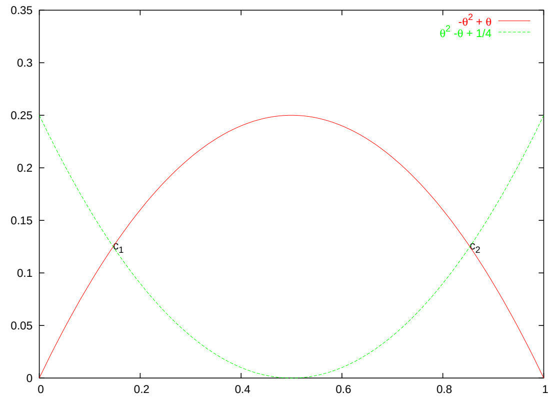

Figura 1: Gráfico em função de θ de $EQM(\hat{\theta}_1)$ e $EQM(\hat{\theta}_2)$

1.10 Sejam $X_1, \ldots, X_n$ uma amostra aleatória de tamanho n da distribuição da varíavel aleatória X com f.d.p. dada por

$$f(x|\theta) = e^{-(x-\theta)}, \qquad x > \theta, \theta > 0$$

(a) Especifique o espaço paramétrico e o suporte associado à distribuição de X.

Do enunciado, temos: $\Theta = \{\theta, \theta > 0\}$ e $A(x) = \{x, x > \theta\}$.

(b) Verifique se $\hat{\theta}_1 = \overline{X}$ e $\hat{\theta}_2 = X_{(1)}$ são estimadores não viciados para θ.

Em primeiro lugar, verifiquemos o valor de $E(X)$.

$$E(X) = \int_{-\infty}^{\infty} x f(x|\theta) \mathrm{d}x = \int_{\theta}^{\infty} x e^{-(x-\theta)} \mathrm{d}x = e^{\theta} \int_{\theta}^{\infty} x e^{-x} \mathrm{d}x \quad (1)$$

Resolvendo essa integral por partes, fazemos $u = x$ e $\mathrm{d}v = e^{-x}\mathrm{d}x$, obtendo $\mathrm{d}u = \mathrm{d}x$ e $v = -e^{-x}$. Pela regra do produto então:

$$(1) = e^{\theta}\left(\left[-xe^{-x}\right]_{\theta}^{\infty} - \int_{\theta}^{\infty} -e^{-x}\mathrm{d}x\right) = e^{\theta}e^{-\theta}(\theta+1) = \theta + 1$$

De imediado segue que $\hat{\theta}_1$ é viciado para θ pois $E(\hat{\theta}_1) = E(\overline{X}) = E(X) = \theta + 1$. Para $\hat{\theta}_2$ precisamos calcular sua esperança manualmente.

Em primeiro lugar, achemos a função de distribuição de $X_{(1)}$:

$$\begin{aligned} F_{X_{(1)}}(x) &= P(X \leq x) = P(\min(X_1,\ldots,X_n) \leq x) \\ &= 1 - P(\min(X_1,\ldots,X_n) > x) \\ &= 1 - P(X_1 > x, \ldots, X_n > x) \\ &= 1 - P(X_1 > x)\ldots P(X_n > x) \end{aligned}$$

Dada a função de densidade de X, é imediata a obtenção de uma expressão para $P(X > a)$:

$$P(X > a) = \int_a^{\infty} e^{-(x-\theta)}\mathrm{d}x = e^{-a+\theta}, \qquad a > \theta > 0$$

Voltando isso na penúltima equação:

$$F_{X_{(1)}}(x) = 1 - e^{-n(x-\theta)} \Rightarrow f_{X_{(1)}}(x) = ne^{-n(x-\theta)}, \qquad x > \theta > 0$$

Temos então:

$$E(\hat{\theta}_2) = E(X_{(1)}) = \int_{\theta}^{\infty} xne^{-n(x-\theta)}\mathrm{d}x = ne^{n\theta}\int_{\theta}^{\infty} xe^{-nx}\mathrm{d}x \qquad (2)$$

Fazendo $u = x$ e $\mathrm{d}v = e^{-nx}\mathrm{d}x$, temos $\mathrm{d}u = \mathrm{d}x$ e $v = -\frac{e^{-nx}}{n}$. Pela regra do produto:

$$(2) = ne^{n\theta}\left(\left[-x\frac{e^{-nx}}{n}\right]_{\theta}^{\infty} - \int_{\theta}^{\infty} -\frac{e^{-nx}}{n}\mathrm{d}x\right) = \frac{n\theta+1}{n} = \theta + \frac{1}{n}$$

Logo, $\hat{\theta}_2$ é um estimador assintoticamente não viciado para θ.

(c) Encontre e compare os EQMs dos dois estimadores. Faça um gráfico como função de θ.

Em primeiro lugar, precisamos calcular $Var(X)$. Para isso, vamos obter $E(X^2)$:

$$E(X^2) = \int_{\theta}^{\infty} x^2 e^{-(x-\theta)}\mathrm{d}x = e^{\theta}\int_{\theta}^{\infty} x^2 e^{-x}\mathrm{d}x \qquad (3)$$

Tomando $u = x^2$ e $\mathrm{d}v = e^{-x}$, temos $\mathrm{d}u = 2x\mathrm{d}x$ e $v = -e^{-x}$. Assim:

$$\begin{aligned} (3) &= e^{\theta}\left(\left[-x^2e^{-x}\right]_{\theta}^{\infty} - \int_{\theta}^{\infty} -e^{-x}2x\mathrm{d}x\right) \\ &= e^{\theta}\left(\theta^2 e^{-\theta} + 2(e^{-\theta}(\theta+1)\right) \\ &= \theta^2 + 2\theta + 2 \end{aligned}$$

Temos que $Var(X) = E(X^2) - E^2(X)$, utilizando os valores já conhecidos:

$$Var(X) = \theta^2 + 2\theta + 2 - (\theta+1)^2 = 1$$

A partir daí podemos calcular $EQM(\hat{\theta}_1)$:

$$EQM(\hat{\theta}_1) = Var(\hat{\theta}_1) + B^2(\hat{\theta}_1) = \frac{Var(X)}{n} + (E(\overline{X}) - \theta)^2 = \frac{1}{n} + 1 = \frac{n+1}{n}$$

Calculemos agora $EQM(\hat{\theta}_2)$. Em primeiro lugar precisamos de $Var(\hat{\theta}_2)$ que precisa de $E(\hat{\theta}_2)$:

$$\begin{aligned}
E(\hat{\theta}_2^2) &= E(X_{(1)}^2) = \int_\theta^\infty x^2 n e^{-n(x-\theta)} \mathrm{d}x = ne^{n\theta} \int_\theta^\infty x^2 e^{-nx} \mathrm{d}x \\
&= ne^{n\theta} \left(\left[\frac{\theta^2 e^{-n\theta}}{n} \right]_\theta^\infty + \frac{2}{n} \int_\theta^\infty x e^{-nx} \mathrm{d}x \right) \\
&= ne^{n\theta} \left(\theta^2 \frac{e^{-n\theta}}{n} + \frac{2}{n} \left(\frac{\theta e^{-n\theta}}{n} + \frac{e^{-n\theta}}{n^2} \right) \right) \\
&= \theta^2 + \frac{2\theta}{n} + \frac{2}{n^2}
\end{aligned}$$

Assim:

$$Var(\hat{\theta}_2) = \theta^2 + \frac{2\theta}{n} + \frac{2}{n^2} - \left(\theta + \frac{1}{n}\right)^2 = \frac{1}{n^2}$$

Calculando então o EQM:

$$EQM(\hat{\theta}_2) = Var(\hat{\theta}_2) + B^2(\hat{\theta}_2) = \frac{1}{n^2} + (E(\hat{\theta}_2) - \theta)^2 = \frac{1}{n^2} + \frac{1}{n^2} = \frac{2}{n^2}$$

O estimador $\hat{\theta}_2$ é melhor que o estimador $\hat{\theta}_1$, pois seu erro quadrático médio é *sempre* menor, para todo θ e fica cada vez menor quanto maior o tamanho da amostra. Não faz sentido comparar os dois através de um gráfico em função de θ pois nenhum dos dois depende de θ.

1.12 Sejam $X_1, \ldots, X_n$ uma amostra aleatória de tamanho n da distribuição de uma variável aleatória $X \sim U(0, \theta)$. Considere os estimadores $\hat{\theta}_1 = c_1\overline{X}$ e $\hat{\theta}_2 = c_2 X_{(n)}$.

(a) Encontre c_1 e c_2 que tornam os estimadores não viciados.
No primeiro caso temos:

$$E(\hat{\theta}_1) = E(c_1\overline{X}) = c_1 E(\overline{X}) = c_1 \frac{\theta}{2}$$

Como queremos que $E(\hat{\theta}_1) = \theta$, fazemos:

$$c_1 \frac{\theta}{2} = \theta \Rightarrow c_1 = 2$$

No segundo, precisamos em primeiro lugar encontrar a esperança de $X_{(n)}$:

$$\begin{aligned}
F_{X_{(n)}}(y) &= P(X_(n) \leq y) = P(\max(X_1, \ldots, X_n)) \\
&= P(X_1 \leq y) \ldots P(X_n \leq y) \\
&= \frac{x^n}{\theta^n} \Rightarrow f_{X_{(n)}}(y) = \frac{nx^{n-1}}{\theta^n}
\end{aligned}$$

Então:

$$E(X_{(n)}) = \int_0^\theta x \frac{nx^{n-1}}{\theta^n} \mathrm{d}x = \frac{n}{\theta^n} \int_0^\theta x^n \mathrm{d}x = \frac{n}{n+1}\theta = \frac{n}{n+1}\theta$$

Como queremos $E(\hat{\theta}_2) = \theta$:

$$c_2 \frac{n}{n+1}\theta = \theta \Rightarrow c_2 = \frac{n+1}{n}$$

(b) Encontre e compare os EQMs dos dois estimadores.

$$\begin{aligned}
EQM(\hat{\theta}_1) &= Var(\hat{\theta}_1) + B^2(\hat{\theta}_1) = Var(2\overline{X}) + (E(\hat{\theta}_1) - \theta)^2 \\
&= \frac{\theta^2}{3n} + 0 = \frac{\theta^2}{3n}
\end{aligned}$$

Como não temos de imediato como acima, vamos encontrar então $Var(\hat{\theta}_2)$. Em primeiro lugar, encontremos $Var(X_{(n)})$, pois a variância que queremos depende dela:

$$Var(X_{(n)}) = E(X_{(n)}^2) - E^2(X_{(n)})$$

Como já obtivemos $E(X_{(n)})$ acima, precisamos apenas encontrar $E(X_{(n)}^2)$:

$$E(X_{(n)}^2) = \int_0^\theta x^2 \frac{nx^{n-1}}{\theta^n} \mathrm{d}x = \frac{n}{\theta^n} \int_0^\theta x^{n+1} \mathrm{d}x = \frac{n}{n+2}\theta^2$$

Assim:

$$Var(X_{(n)}) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \theta^2 \left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right)$$

Calculando então $Var(\hat{\theta}_2)$:

$$\begin{aligned}
Var(\hat{\theta}_2) &= Var\left(\frac{n+1}{n} X_{(n)}\right) = \left(\frac{n+1}{n}\right)^2 Var(X_{(n)}) \\
&= \left(\frac{n+1}{n}\right)^2 \theta^2 \left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right) \\
&= \theta^2 \left(\frac{n^2+2n+1}{n^2+2n} - 1\right) = EQM(\hat{\theta}_2)
\end{aligned}$$

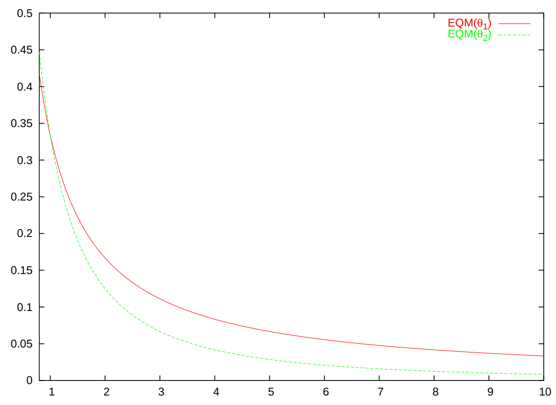

Figura 2: $EQMs$ em função de n

Na Figura 2 temos o gráfico dos $EQMs$ de acordo com o tamanho da amostra n. Fica claro que o erro quadrático médio do segundo estimador é sempre menor ou igual (com menor ou igual substituido por estritamente menor para qualquer $n > 1$) que o erro quadrático médio do primeiro estimador. Logo o estimador $\hat{\theta}_2$ é um melhor estimador para θ que $\hat{\theta}_1$.

1.13 Sejam $X_1, \ldots, X_n$ uma amostra aleatória de tamanho n da distribuição da variável aleatória $X \sim N(0, \sigma^2)$. Seja $S^2 = \sum_{i=1}^n X_i^2$. Considere os estimadores

$$\hat{\sigma}^2 = cS^2$$

(a) Encontre o EQM do estimador acima.

Em primeiro lugar, precisamos achar a esperança e a variância de S^2:

$$E(S^2) = E\left(\sum_{i=1}^n X_i^2\right) = E(X_1^2) + \ldots + E(X_n^2)$$

Notemos que:

$$Var(X) = E(X^2) - E^2(X) \Rightarrow E(X^2) = Var(X) + E^2(X) = \sigma^2$$

Voltando esse resultado na equação acima:

$$E(S^2) = n\sigma^2$$

Para achar a variância de S^2 resta-nos apenas encontrar $E((S^2)^2)$. Para isto, notemos que $\sum_{i=1}^n \frac{X_i - \mu}{\sigma^2}$ segue uma distribuição de qui-quadrado com n graus de liberdade. No caso presente, notando que $\mu = 0$, temos que S^2/σ^2 segue uma distribuição qui-quadrado com n graus de liberdade.

Notemos ainda que a função geradora de momentos de uma variável aleatória que segue uma distribuição de qui-quadrado com n graus de liberdade é dada por:

$$\phi(t) = \left(\frac{1}{1-2t}\right)^{n/2}, \qquad \text{para} \quad t < \frac{1}{2} \tag{4}$$

É imediato que se S^2/σ^2 segue uma distribuição de qui-quadrado com n graus de liberdade, $(S^2/\sigma^2)^2 = (S^2)^2/\sigma^4$ segue uma distribuição qui-quadrado *ao quadrado*, com os mesmos graus de liberdade. Notemos entretanto que estamos interessados justamente na *esperança* de $(S^2)^2$. Podemos obter facilmente através do segundo momento da distribuição qui-quadrado, o valor de $(S^2)^2/\sigma^4$. Basta então notarmos que:

$$E((\chi_n^2)^2) = E(S^2)/\sigma^4 \Rightarrow E((S^2)^2) = E((\chi_n^2)^2)\sigma^4$$

Resta somente então obtermos o valor de $E((\chi_n^2)^2$, que nada mais é que $\phi''(0)$.

$$\begin{aligned} \phi''(t) &= \left(-\left(-1+2\,t\right)^{-1}\right)^{1/2\,n} n\,(n+2)\,(-1+2\,t)^{-2} \\ &\Rightarrow \quad \phi''(0) = n^2 + 2n \end{aligned}$$

Temos então que $E((S^2)^2) = (n^2 + 2n)\sigma^4$. E então:

$$Var(S^2) = (n^2 + 2n)\sigma^4 - (n\sigma^2)^2 = 2n\sigma^4$$

Calculando o EQM:

$$\begin{aligned} EQM(\hat{\sigma}^2) &= EQM(cS^2) = Var(cS^2) + B^2(cS^2) \\ &= c^2 Var(S^2) + (E(cS^2) - \sigma^2)^2 \\ &= c^2 Var(S^2) + (cE(S^2) - \sigma^2)^2 \\ &= 2\,c^2 n\sigma^4 + c^2 n^2 \sigma^4 - 2\,cn\sigma^4 + \sigma^4 \end{aligned}$$

(b) Encontre o valor de c que minimiza o EQM em (a).
Tomemos como função de uma variável:

$$f(c) = 2\,c^2 n\sigma^4 + c^2 n^2 \sigma^4 - 2\,cn\sigma^4 + \sigma^4$$

Fazendo $f'(c) = 0$ obtemos como candidato para máximo ou mínimo local $c = \frac{1}{n+2}$. Para verificarmos se esse ponto é de mínimo ou máximo, tomemos:

$$f''(c) = 4\,n\sigma^4 + 2\,n^2\sigma^4$$

Como $f''(c)$ é sempre positiva, pois n é positivo, temos que $\frac{1}{n+2}$ é um ponto de mínimo local. Logo $c = \frac{1}{n+2}$ minimiza o erro quadrático médio desse estimador com relação a σ^2.

Sobre

A versão eletrônica desse arquivo pode ser obtida em `http://www.feferraz.net`