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

1. Propriedades dos Estimadores
   - Viés
   - Variância de um Estimador
   - Erro Quadrático Médio
   - Estimador Não Viesado de Variância Mínima Uniformemente

Introdução

Propriedades dos estimadores são uma questão fundamental na inferência estatística, pois elas determinam a qualidade e a confiabilidade das estimativas obtidas a partir dos dados amostrais. Um estimador é uma função dos dados amostrais que é usada para estimar um parâmetro desconhecido da população subjacente. As propriedades dos estimadores descrevem como eles se comportam em diferentes situações, como a amostra aumentando de tamanho ou a população apresentando diferentes características. As principais propriedades dos estimadores incluem viés, consistência, eficiência e robustez. O conhecimento dessas propriedades é essencial para avaliar a qualidade de um estimador e para tomar decisões informadas com base nas estimativas obtidas a partir dos dados.

Ao criar um estimador de parâmetros, uma questão fundamental é saber se o estimador difere ou não do parâmetro de maneira sistemática.

Definição 1

Seja $X = (X_1, X_2, ..., X_n)$ *uma amostra aleatória de uma variável aleatória* X *com função de densidade* $f(x;\theta)$, $\theta \in \Omega$. *Seja* $T = T(X_1, X_2, \ldots, X_n)$ *um Estimador de* θ. *O* ***Viés*** *(*$b(\theta)$*) é a média da diferença de* $T(X) - \theta$*, isto é,*

$$b(T(X)) = E(T(X)) - \theta$$

.

Definição 2

Seja $X = (X_1, X_2, ..., X_n)$ *uma amostra aleatória de uma variável aleatória* X *com função de densidade* $f(x;\theta)$, $\theta \in \Omega$. *Seja* $T = T(X_1, X_2, \ldots, X_n)$ *um Estimador de* θ. *Dizemos que* T *é um estimador não viesado de* θ *se* $E(T) = \theta$, *ou seja, se*

$$b(T(X)) = 0.$$

Definição 2

Seja $X = (X_1, X_2, ..., X_n)$ *uma amostra aleatória de uma variável aleatória* X *com função de densidade* $f(x;\theta)$, $\theta \in \Omega$. *Seja* $T = T(X_1, X_2, \ldots, X_n)$ *um Estimador de* θ. *Dizemos que* T *é um estimador não viesado de* θ *se* $E(T) = \theta$, *ou seja, se*

$$b(T(X)) = 0.$$

Definição 3

Se $\lim_{n \to \infty} b(T(X)) = 0$ *para todo* $\theta \in \Theta$, *dizemos que o estimador* $T(X)$ *é assintoticamente não viciado para* θ.

Exemplos 1

1. **Média amostral**: A média amostral é um estimador não viesado da média populacional. Isso significa que, em média, a média amostral estima corretamente a média populacional, sem inclinação para superestimá-la ou subestimá-la.
2. **Variância amostral**: A variância amostral é um estimador não viesado da variância populacional. Isso significa que, em média, a variância amostral estima corretamente a variância populacional, sem inclinação para superestimá-la ou subestimá-la.
3. **Estimador de máxima verossimilhança**: O estimador de máxima verossimilhança é um estimador não viesado que utiliza a função de verossimilhança para encontrar o valor mais provável do parâmetro populacional. Ele é amplamente utilizado em estatística inferencial para estimar parâmetros de distribuições populacionais.

Exemplos 2

4. **Mediana amostral**: A mediana amostral é um estimador não viesado da mediana populacional. Ao contrário da média, que pode ser influenciada por valores extremos, a mediana é mais robusta a essas observações e fornece uma medida mais representativa do centro da distribuição.
5. **Estimador por momentos**: O estimador por momentos é um estimador não viesado que utiliza momentos da amostra para estimar parâmetros populacionais. Ele é uma abordagem simples e amplamente utilizada para estimar média, variância e outros momentos de uma distribuição populacional.
6. **Desvio padrão amostral**: O desvio padrão amostral é um estimador viesado da variância populacional. Ele tende a subestimar a variância populacional, especialmente em amostras pequenas.

Exemplos 3

7. **Média amostral truncada**: Este é um estimador de uma média populacional que é obtido a partir de uma amostra, mas em que alguns valores extremos são removidos antes de se calcular a média amostral. Se a amostra não for representativa da população, a média amostral truncada pode ser um estimador viesado.
8. **Máximo da amostra**: O máximo da amostra é um estimador da ordem estatística mais alta da população. No entanto, se a amostra não for grande o suficiente, o máximo da amostra pode ser um estimador viesado.
9. **Variância amostral não-corrigida**: A variância amostral é um estimador da variância populacional, mas se não for corrigida, isto é, dividida pelo tamanho da amostra menos 1, ela pode ser um estimador viesado.

Exemplo 4

Suponha que a função densidade de probabilidade de uma amostra aleatória $X = (X_1, X_2, ..., X_n)$ seja $\Gamma(1, \theta)$, isto é, $f(x) = \theta^{-1} \exp\left(\frac{-x}{\theta}\right)$, com suporte $0 < x < \infty$. Nesse caso, a distribuição gamma é também chamada de distribuição exponencial. O log da função de verossimilhança é dado por:

$$\ell(\theta) = \log \prod_{i=1}^{n} \frac{1}{\theta} e^{-x_i/\theta} = -n \log \theta - \theta^{-1} \sum_{i=1}^{n} x_i$$

A primeira derivada parcial do log-verossimilhança com respeito a θ é:

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

Continuação do Exemplo 4

Definindo esta derivada parcial como 0 e resolvendo para θ, obtemos a solução $\bar{x}$. Há apenas um valor crítico e, além disso, a segunda derivada parcial do logaritmo da verossimilhança avaliada em $\bar{x}$ é estritamente negativa, o que confirma que ela fornece um máximo. Portanto, para este exemplo, a estatística $\hat{\theta} = \bar{X}$ é a estimativa de máxima verossimilhança (MLE) de θ. Como $E(X) = \theta$, temos que $E(\bar{X}) = \theta$ e, portanto, $\hat{\theta}$ é um estimador não enviesado de θ.

Exemplo 5

Seja $X = (X_1, X_2, ..., X_n)$ ensaios de Bernoulli com parâmetro de sucesso p, definimos o estimador para p como sendo $d(X) = \bar{X}$, a média amostral. Então,

$$\begin{aligned} E_p(\bar{X}) &= \frac{1}{n}(E(X_1) + \cdots + E(X_n)) \\ &= \frac{1}{n}(p + \cdots + p) = p \end{aligned}$$

Assim, $\bar{X}$ é um estimador não viesado para p. Neste caso, geralmente escrevemos $\hat{p}$ em vez de $\bar{X}$, para representar a média amostral.

Continuação do Exemplo 5

Podemos usar o fato de que, para variáveis aleatórias independentes, a variância da soma é a soma das variâncias, assim:

$$Var(\hat{p}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{p(1-p)}{n}$$

Exemplo 6

Se $X_1, \ldots, X_n$ formarem uma amostra aleatória simples com média finita e desconhecida μ, então $\overline{X}$ é um estimador não enviesado de μ. Se os X_i possuem variância σ^2, então:

$$Var(\overline{X}) = \frac{\sigma^2}{n}$$

Continuação do Exemplo 6

Com relação à variância amostral, temos

$$E[\hat{\sigma}^2] = \frac{1}{n}E\left[\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = \frac{1}{n}\sum_{i=1}^{n}E[(X_i - \overline{X})^2]$$
$$= \frac{1}{n}\sum_{i=1}^{n}E\left[\left[(X_i - \mu) - (\overline{X} - \mu)\right]^2\right] = \frac{(n-1)}{n}\sigma^2$$

Portanto, $\hat{\sigma}^2$ é viciado para σ^2, mas é assintoticamente não viciado, ou seja, à medida que o tamanho da amostra aumenta, o viés diminui.

Variância de um Estimador

É desejável que os estimadores de um determinado parâmetro da população possuam um valor de variância que seja o mínimo possível, isso porque uma variância baixa significa uma precisão maior da estimativa do que uma variância alta.

Podemos observar na Figura (Colocar Figura) que o histograma das médias aritméticas possui uma variabilidade menor do que o histograma dos primeiros valores das amostras. Além disso, ao aumentarmos o tamanho de cada amostra a variabilidade da média aritmética da amostra vai diminuindo, enquanto que a variabilidade do primeiro valor da amostra não diminui à medida que o tamanho da amostra aumenta.

Pode-se mostrar que a média aritmética da amostra é o estimador de menor variância entre todos os estimadores lineares da média de uma população. **O erro-padrão de um estimador é a raiz quadrada da variância do estimador.**

Erro Quadrático Médio

Podemos avaliar a qualidade de um estimador calculando seu erro quadrático médio, definido por:

$$E[(T(X) - \theta)^2] \tag{1}$$

Estimadores com erro quadrático médio menor são geralmente preferidos em relação a aqueles com erro quadrático médio maior.

Se escrevermos $Y = T(X) - \theta$ em (1) e lembrarmos que a variância é dada por $\text{Var}(Y) = E(Y^2) - (E(Y))^2$, então:

$$
\begin{aligned}
E(Y) &= E(T(X) - \theta) \\
&= E(T(X)) - \theta \\
&= b(T(X))
\end{aligned}
$$

e,

$$\text{Var}(Y) = \text{Var}(T(X))$$

Dessa forma, o erro quadrático médio é dado por:

$$
\begin{aligned}
E[(T(X) - \theta)^2] = E(Y^2) &= \text{Var}(Y) + (E(Y))^2 \\
&= \text{Var}(T(X)) + b^2(T(X))
\end{aligned}
$$

Assim, a representação do erro quadrático médio como igual à variância do estimador mais o quadrado do viés é chamada de decomposição viés-variância. O erro quadrático médio pode ser considerado como uma medida da precisão de um estimador. Se a variância for pequena, podemos dizer que o estimador é preciso. Ele ainda pode não ser muito preciso se o viés for grande. Observem que:

- O erro quadrático médio para um estimador não viciado é a sua variância.

Assim, a representação do erro quadrático médio como igual à variância do estimador mais o quadrado do viés é chamada de decomposição viés-variância. O erro quadrático médio pode ser considerado como uma medida da precisão de um estimador. Se a variância for pequena, podemos dizer que o estimador é preciso. Ele ainda pode não ser muito preciso se o viés for grande. Observem que:

- O erro quadrático médio para um estimador não viciado é a sua variância.
- O viés sempre aumenta o erro quadrático médio.

O erro quadrático médio é comumente empregado na comparação de estimadores. Dizemos que o estimador $\hat{\theta}_1$ é melhor que o estimador $\hat{\theta}_2$ se

$$EQM[\hat{\theta}_1] \leq EQM[\hat{\theta}_2], \tag{2}$$

para todo θ, com $\leq$ substituído por $<$ pelo menos para um valor de θ. Nesse caso, o estimador $\hat{\theta}_2$ é dito ser inadmissível. Se existir um estimador $\hat{\theta}^*$ tal que para todo estimador $\hat{\theta}$ de θ com $\hat{\theta} \neq \hat{\theta}^*$

$$EQM[\hat{\theta}^*] \leq EQM[\hat{\theta}], \tag{3}$$

para todo θ com $\leq$ substituído por $<$ para pelo menos um θ, então $\hat{\theta}^*$ é dito ser ótimo para θ.

Notemos que, se em (3) os estimadores são não viciados, então $\hat{\theta}^*$ é dito ser o estimador não viciado de variância uniformemente mínima, se

$$Var[\hat{\theta}^*] \leq Var[\hat{\theta}], \tag{4}$$

para todo θ, com $\leq$ substituído por $<$ para pelo menos um θ.

Exemplo 1

Sejam X_1, X_2, X_3 uma amostra aleatória da variável aleatória X com $E[X] = \theta$ e $\mathrm{Var}[X] = 1$. Consideremos os estimadores

$$\hat{\theta}_1 = \bar{X} = \frac{X_1 + X_2 + X_3}{3} \text{ e } \hat{\theta}_2 = \frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{4}X_3.$$

Continuação do Exemplo 1

Temos que,

$$E[\hat{\theta}_1] = \theta \quad \text{e} \quad Var[\hat{\theta}_1] = \frac{1}{3}$$
$$E[\hat{\theta}_2] = \theta \quad \text{e} \quad Var[\hat{\theta}_2] = \frac{6}{16}$$

Como ambos $\hat{\theta}_1$ e $\hat{\theta}_2$ são não viesados, segue que $\bar{X}$ é um estimador melhor que $\hat{\theta}_2$, pois $Var[\bar{X}] < Var[\hat{\theta}_2]$ para todo θ.

Exemplo 2

Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória $X \sim N(\mu, \sigma^2)$. Conforme visto anteriormente, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ é um estimador viciado para σ^2. Sabe-se

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

é um estimador não viciado para σ^2. Por outro lado, temos que

$$EQM[S^2] = Var[S^2] = \frac{2\sigma^4}{n-1}, \quad EQM[\hat{\sigma}^2] = \frac{2\sigma^4}{n-1}\left(1 - \frac{3n-1}{2n^2}\right).$$

Continuação do Exemplo 2

Notemos que $\hat{\sigma}^2$, apesar de viciado, apresenta um EQM menor que o EQM do estimador S^2.

Exemplo 3

Sejam $X_1, \ldots, X_n$ uma amostra aleatória de tamanho n da variável aleatória X, com distribuição de Bernoulli com parâmetro θ, ou seja, $Binomial(1, \theta)$. Conforme visto no modelo binomial, $Y = X_1 + \ldots + X_n$ tem distribuição binomial $Binomial(n, \theta)$. Consideremos os estimadores

$$\hat{\theta}_1 = X = \frac{Y}{n} \text{ e } \hat{\theta}_2 = \frac{(Y + \sqrt{n}/2)}{(n + \sqrt{n})}$$

Continuação do Exemplo 3

Como $E[\bar{X}] = \theta$, temos que

$$EQM[\hat{\theta}_1] = Var[\bar{X}] = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n}.$$

Por outro lado,

$$\begin{aligned} E[\hat{\theta}_2] &= E\left[\frac{Y + \sqrt{n}/2}{n + \sqrt{n}}\right] \\ &= \frac{n\theta + \sqrt{n}/2}{n + \sqrt{n}} \\ &= \frac{n}{n + \sqrt{n}} \cdot \theta + \frac{\sqrt{n}/2}{n + \sqrt{n}}. \end{aligned}$$

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

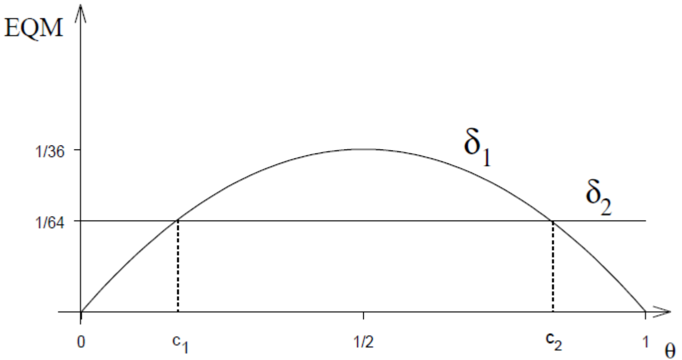
Continuação do Exemplo 3

Notemos que, na verdade, o vício é uma função linear de θ. Portanto,

$$\begin{aligned} EQM[\hat{\theta}_2] &= E[(\hat{\theta}_2 - \theta)^2] \\ &= \frac{1}{(n+\sqrt{n})^2}\left[Var[Y] + n\left(\frac{1}{2} - \theta\right)^2\right] = \frac{n}{4(n+\sqrt{n})^2} \end{aligned}$$

Um fato importante a ser notado é que o EQM do estimador $\hat{\theta}_2$ é independente de θ. O EQM dos dois estimadores é representado graficamente na Figura do próximo slide, para $n = 9$.

Figura 1.1. EQM de $\delta_1 = \hat{\theta}_1$ e $\delta_2 = \hat{\theta}_2$



Continuação do Exemplo 3

Temos, então, que nenhum dos estimadores é melhor uniformemente, isto é, para todo θ. Para $c_1 < \theta < c_2$, $EQM[\hat{\theta}_2] < EQM[\hat{\theta}_1]$, ou seja, $\hat{\theta}_2$ é melhor que $\hat{\theta}_1$. Por outro lado, para $\theta < c_1$ ou $\theta > c_2$, temos que $EQM[\hat{\theta}_1] < EQM[\hat{\theta}_2]$, ou seja, $\hat{\theta}_1$ é melhor que $\hat{\theta}_2$.

Estimador Não Viesado de Variância Mínima Uniformemente (ENVVMU)

Um estimador $\hat{\theta}^*$ é um estimador não viesado de θ se satisfaz $E(\hat{\theta}^*) = \theta$ e, para qualquer outro estimador $\hat{\theta}$, temos que $Var(\hat{\theta}^*) \leq Var(\hat{\theta})$, $\forall\ \theta$. $\hat{\theta}^*$ é também chamado de Estimador Não Viesado de Variância Mínima Uniformemente (ENVVMU) de θ.

Eficiência Relativa de um Estimador

O erro quadrático médio é uma métrica importante e pode ser usado para definir a eficiência relativa de um estimador comparado a outro:

$$EFR(\hat{\theta}_1, \hat{\theta}_2) = \frac{EQM(\hat{\theta}_1)}{EQM(\hat{\theta}_2)}$$

Se $E(\hat{\theta}_1, \hat{\theta}_2) < 1$, conclui-se que $\hat{\theta}_1$ é um estimador superior a $\hat{\theta}_2$ ou vice-versa.

Referências I