Estimadores Baseados em Estatística Suficiente

Professor: Pedro M. Almeida-Junior

22 de fevereiro de 2022

Departamento de Estatística (UEPB)

Estimadores baseados em estatísticas suficientes

Seja $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória X com função de densidade (ou de probabilidade) $f(x, \theta)$. Seja $T = T(X_1, \ldots, X_n)$ uma estatística suficiente para θ e $S = S(X_1, \ldots, X_n)$ um estimador de θ que **não é uma função da estatística suficiente** T. Então,

$$\hat{\theta} = \mathbb{E}\left(S \mid T\right),$$

é um **estimador** de θ, ou seja, é uma função de T que não depende de θ, pois, sendo T suficiente, a distribuição condicional de $X_1, \ldots, X_n$ dado T é independente de θ e $S(X_1, \ldots, X_n)$ é uma função que depende apenas de $X_1, \ldots, X_n$.

Ainda se S é um estimador não viesado de θ, então $\hat{\theta}$ é também não viesado para θ. Pois,

$$\mathbb{E}\left(\hat{\theta}\right) = \mathbb{E}\left[\mathbb{E}\left(S \mid T\right)\right] = \mathbb{E}(S) = \theta$$

(Como S não depende de T, logo, $\mathbb{E}\left[S \mid T\right] = \mathbb{E}(S)$)

(Como S é não viesado, então, $\mathbb{E}\left[S\right] = \theta$)

- Muitos estimadores são baseados em estatísticas suficientes. Um importante resultado que faz essa relação entre estimadores não viesados e estatísticas suficientes é o chamado **Teorema de Rao-Blackwell**.

Teorema de Rao-Blackwell: Sejam $X_1, \ldots, X_n$ uma amostra aleatória e $T = T(X_1, \ldots, X_n)$ uma estatística suficiente para um parâmetro θ. Considere $S = S(X_1, \ldots, X_n)$ um estimador não viesado de θ, com $\mathbb{E}(S) < \infty$, para todo $\theta \in \Theta$. Então, existe um estimador

$$\hat{\theta} = \mathbb{E}\left[S \mid T\right]$$

tal que

$$\mathrm{Var}[\hat{\theta}] \leq \mathrm{Var}[S],$$

para todo θ.

Portanto, temos que o estimador $\hat{\theta}$ baseado na estatística suficiente T apresenta uma variância menor (ou igual) que a variância do estimador não viciado S. Desse modo, **qualquer estimador S que não é função de uma estatística suficiente pode ser melhorado.**

Estatística Completa

Agora vamos apresentar o conceito de **estatística completa**, que em conjunto com a definição de **suficiência**, possibilita encontrar um estimador ótimo, isto é, o **estimador não viesado de variância uniformemente mínima (ENVVUM)**.

Definição: Uma estatística $T = T(X_1, \ldots, X_n)$ é dita ser completa em relação à família $f(x \mid \theta) : \theta \in \Theta$, se a única função real g, definida no domínio de T, tal que $E[g(T)] = 0$, para todo θ é a função nula, isto é, $g(T) = 0$ com probabilidade 1.

Exemplo: Considerando uma amostra com distribuição binomial, assumindo $T = \sum_{i=1}^{n} X_i$. Temos que

$$E[g(T)] = \sum_{x=0}^{n} g(x) \left(\begin{array}{c} n \\ x \end{array} \right) \theta^x (1-\theta)^{n-x} = 0 \text{ para todo } \theta,$$

se e somente se

$$\sum_{x=0}^{n} g(x) \left(\begin{array}{c} n \\ x \end{array} \right) \rho^x = 0, \quad \text{para todo} \quad \rho$$

onde $\rho = \theta/(1-\theta)$. Como o lado esquerdo de (2.5.3) é um polinômio em ρ de grau n temos que $g(x) = 0$ para todo x. Portanto $T = \sum_{i=1}^{n} X_i$ é completa em relação à família Binomial.

Estatística suficiente e completa (família exponencial)

Teorema: Suponha que X tenha distribuição da família exponencial k dimensional. Então, a estatística

$$T(\mathrm{X}) = \left(\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i) \right)$$

é suficiente para θ. $T(\mathrm{X})$ será também completa desde que o domínio de variação de $(c_1(\theta), \ldots, c_k(\theta))$ contenha um retângulo k-dimensional.

Em resumo, a estatística k-Dimensional $T(X) = (T_1(x), \ldots, T_k(x))$ será **completa** desde que o espaço paramétrico Θ contenha um conjunto aberto em $\Re^k$, ou seja, **a dimensão do espaço paramétrico não pode ser menor que a dimensão de $C(\theta)$.**

Estimadores ENVVUM

Teorema (Lehmann-Scheffé): Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória X com f.d.p. (ou f.p.), $f(x \mid \theta)$. Seja T uma estatística suficiente e completa. Seja S um estimador não viciado de θ. Então $\hat{\theta} = E(S \mid T)$ é o único estimador não viciado de θ baseado em T e **é o estimador não viciado de variância uniformemente mínima (ENVVUM)** para θ.

- Se T é uma estatística suficiente e a família de distribuições de T é completa, então, T é uma **estatística suficiente e completa**.

- No caso em que temos uma estatística suficiente e completa T, se obtivermos $\varphi(T)$, estimador não viesado de θ, este será o **estimador não viesado de mínima variância** de θ.

Exemplo: Sejam $X_1, \ldots, X_n$ uma amostra aleatória da distribuição Poisson com parâmetro θ. Mostre que o estimador $\bar{X}$ é ENVVUM.