Estatística não paramétrica

Fernando Lucambio

Departamento de Estatística
Universidade Federal do Paraná

6 de março de 2024

Sejam $X_1, X_2, \cdots, X_n$ variáveis aleatórias independentes identicamente distribuídas com distribuição comum $\mathcal{L}(X)$ e seja $\mathcal{P}$ a classe de todas as possíveis distribuições de X que consiste de todas as distribuições absolutamente contínuas ou discretas.

Definição:

A estatística $T(X)$ é suficiente para a família de distribuições $\mathcal{P}$ se a distribuição condicional de $X|T = t$ é a mesma, seja qual for a função de distribuição $F \in \mathcal{P}$.

Exemplo:

Sejam $X_1, X_2, \cdots, X_n$ variáveis aleatórias independentes igualmente distribuídas com distribuição absolutamente contínua e seja $T = (X_{(1)}, \cdots, X_{(n)})$ a estatística de ordem. Então

$$f(x|T = t) = \frac{1}{n!},$$

e vemos que T é uma estatística suficiente para a família das distribuições absolutamente contínuas.

Definição:

A família de distribuições $\mathcal{P}$ é completa se somente a função zero for o estimador não viesado de 0, isto é, $\mathrm{E}_F\Big(h(X)\Big) = 0$, para todo $F \in \mathcal{P}$ implica que $h(X) = 0$.

Teorema:

A estatística de ordem $(X_{(1)}, \cdots, X_{(n)})$ é uma estatística suficiente e completa desde que a amostra $X_1, X_2, \cdots, X_n$ seja composta de variáveis aleatórias independentes identicamente distribuídas do tipo discreta ou contínua.

Definição:

Diz-se que uma função real $g(F)$ é estimável se tiver um estimador não viciado, isto é, se existe uma estatística $T(X)$ tal que

$$\mathrm{E}_F\Big(T(X)\Big) = g(F),$$

para todo $F \in \mathcal{P}$.

Exemplo:

Se $\mathcal{P}$ é a classe de todas as distribuições para as quais o segundo momento existe, X é um estimador não viciado de $\mu(F)$, a média da população. Similarmente

$$\mu_2(F) = \mathrm{Var}_F(X),$$

é também estimável e um estimador não viciado é

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

Da mesma forma, $\overline{X} - \overline{Y}$ é um estimador não viciado de $\mathrm{E}(X) - \mathrm{E}(Y)$,

$$T(X_1, \cdots, X_n) = \frac{1}{n}\{\text{número de } X > c\}$$

é um estimador não viciado de $P_F(X > c)$ e assim por diante.

Definição:

O grau m, $m \geq 1$, de um parâmetro estimável $g(F)$ é o menor tamanho de amostra para o qual o parâmetro é estimável, ou seja, é o menor n para o qual existe um estimador não viciado $T(X_1, \cdots, X_n)$ com

$$\mathrm{E}_F\Big(T(X)\Big) = g(F),$$

para todo $F \in \mathcal{P}$.

Exemplo:

O parâmetro $g(F) = P_F(X > c)$, c conhecida, têm grau 1. Também $\mu(F)$ é estimável com grau 1 assumindo que existe ao menos um $F \in \mathcal{P}$ tal que $\mu(F) \neq 0$. $\mu_2(F)$ é estimável com grau 2, ou seja, ao menos duas observações são necessárias. De maneira similar, $\mu^2(F)$ têm grau 2.

Definição:

Um estimador não viciado de algum parâmetro baseado no tamanho mínimo de amostra, ou seja, com amostra igual ao grau m é chamado de kernel.

Exemplo:

Seja $X_1, \cdots, X_n$ uma amostra aleatória com distribuição F. Então X_i é o kernel de $\mu(F)$; X_iX_j, $i \neq j$, é o kernel de $\mu^2(F)$ e cada

$$T(X_i, X_j) = X_i^2 - X_iX_j, \qquad i = 1, \cdots, n, \quad i \neq j,$$

é kernel de $\mu_2(F)$.

Teorema:

Existe um kernel simétrico para cada parâmetro estimável.

Demonstração

Seja $T(X_1, \cdots, X_m)$ um kernel para $g(F)$. Também é

$$T_s(X_1, \cdots, X_m) = \frac{1}{m!} \sum_P T(X_{i_1}, X_{i_2} \cdots, X_{i_m})$$

um kernel para $g(F)$, onde a soma P acontece sobre todas as $m!$ permutações de $\{1, 2, \cdots, m\}$.

Exemplo:

Seja $X_1, \cdots, X_n$ uma amostra aleatória com distribuição F. Um kernel simétrico para $\mu_2(F)$ é

$$\begin{aligned} T_s(X_i, X_j) &= \frac{1}{2}\Big(T(X_i, X_j) + T(X_j, X_i)\Big) \\ &= \frac{1}{2}(X_i^2 - X_i X_j) + \frac{1}{2}(X_j^2 - X_j X_i) \\ &= \frac{1}{2}(X_i - X_j)^2, \end{aligned}$$

para $i = 1, \cdots, n$ e $i \neq j$.

Definição:

Seja $g(F)$ um parâmetro estimável de grau m e $X_1, X_2, \cdots, X_n$ uma amostra aleatória de F de tamanho n, $n \geq m$. Correspondendo a qualquer kernel $T(X_1, \cdots, X_n)$ de $g(F)$, definimos a U-estatística para a amostra como

$$U(X_1, X_2, \cdots, X_n) = \frac{1}{\binom{n}{m}} \sum_{C} T_s(X_{i_1}, X_{i_2}, \cdots, X_{i_m}),$$

onde o índice da soma C percorre todas as $\binom{n}{m}$ permutações de m inteiros $(i_1, i_2, \cdots, i_m)$ escolhidos de $\{1, 2, \cdots, n\}$ e T_s é um kernel simétrico, como definido na demonstração do Teorema anterior.

Claramente, a U-estatística definida é simétrica nos X e

$$\mathrm{E}_F\big(U(X)\big) = g(F),$$

para todo F.

Exemplo:

Seja $X_1, \cdots, X_n \sim F$. Para estimarmos $\mu(F)$ a U-estatística é dada por

$$U(X_1, X_2, \cdots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Para estimarmos $\mu_2(F)$, um kernel simétrico é

$$T_s(X_{i_1}, X_{i_2}) = \frac{1}{2}\big(X_{i_1} - X_{i_2}\big)^2,$$

para $i_1 = 1, 2, \cdots, n$, $i_1 \neq i_2$.

Exemplo (continuação)

A correspondente U-estatística é

$$U(X_1, X_2, \cdots, X_n) = \frac{1}{\binom{n}{2}} \sum_{i_1 < i_2} \frac{1}{2} (X_{i_1} - X_{i_2})^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 = S^2.$$

De maneira similar, para estimarmos $\mu_2(F) = \text{Var}_F(X)$,, o kernel simétrico é $T_s(X_{i_1}, X_{i_2}) = X_{i_1} X_{i_2}$ e a correspondente U-estatística é

$$U(X) = \frac{1}{\binom{n}{2}} \sum_{i<j} X_i X_j = \frac{1}{n(n-1)} \sum_{i \neq j} X_i X_j.$$

Para estimarmos $\mu^3(F)$, ou seja, o cubo da esperança populacional, um kernel simétrico seria $T_s(X_{i_1}, X_{i_2}, X_{i_3}) = X_{i_1}X_{i_2}X_{i_3}$. Observe que

$$\begin{aligned} T_s(X_{i_1}, X_{i_2}, X_{i_3}) &= X_{i_1}X_{i_2}X_{i_3} \\ T_s(X_{i_2}, X_{i_1}, X_{i_3}) &= X_{i_2}X_{i_1}X_{i_3} = X_{i_1}X_{i_2}X_{i_3} \\ T_s(X_{i_3}, X_{i_2}, X_{i_1}) &= X_{i_3}X_{i_1}X_{i_1} = X_{i_1}X_{i_2}X_{i_3}. \end{aligned}$$

Sendo que a U-estatística é

$$U(X) = \frac{1}{\binom{n}{3}} \sum_{i<j<k} X_iX_jX_k = \frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} X_iX_jX_k.$$

O seguinte resultado mostra a importância da U-estatística.

Teorema:

Seja $\mathcal{P}$ a classe de todas as distribuições absolutamente contínuas ou discretas. Qualquer função estimável $g(F)$, $F \in \mathcal{P}$, tem um estimador único que é não viciado, simétrico nas observações e uniformemente de variância mínima entre todos os estimadores não viciados.

Teorema:

Seja $T(X_1, \cdots, X_n)$ um estimador não viciado para $g(F)$, $F \in \mathcal{P}$. A correspondente U-estatística é essencialmente o único estimador não viciado uniformemente de mínima variância.

De acordo com os teoremas anteriores precisamos apenas considerar estimadores que sejam simétricos nas observações e tudo o que devemos fazer é torná-las não viciados.

Este procedimento leva a um estimador não viciado com a menor variância na classe de todos os estimadores não viciados do parâmetro.

Por exemplo, como consequência destes teoremas, $\overline{X}$ e S^2 são os únicos estimadores não viciados uniformemente de variância mínima de $\mu(F)$ e $\mu_2(F)$, respectivamente.

Exemplo:

Seja $\mathcal{P}$ a classe de todas as distribuições absolutamente contínuas e $X_1, X_2, \cdots, X_n$ uma amostra aleatória de tamanho n. Para estimarmos

$$g(F) = P_F(X_1 > c),$$

onde c é uma constante fixa, definimos

$$Y_i = \begin{cases} 1, & X_i > c, \\ 0, & X_i \leq c \end{cases} \qquad i = 1, 2, \cdots, n\cdot$$

Considere agora $T(Y_1, Y_2, \cdots, Y_n) = \sum_{i=1}^{n} \alpha_i Y_i$, como um estimador de $g(F)$. Para encontrar o estimador não viciado de mínima variância de g simetrizamos T nos $Y_1, Y_2, \cdots, Y_n$.

Exemplo (continuação):

Isso acontece se $\alpha_i = \alpha$, $\forall i$ e $T(Y) = \alpha \sum_{i=1}^{n} Y_i$. Para T ser não viciado, temos que

$$\mathrm{E}_F(T) = \alpha \sum_{i=1}^{n} \mathrm{E}_F(Y_i) = \alpha n g(F) = g(F), \quad \text{se} \quad \alpha = \frac{1}{n}.$$

Portanto, $\frac{1}{n}\sum_{i=1}^{n} Y_i$ é o estimador não viciado de mínima variância; também

$$\mathrm{Var}_F(T) = \frac{g(F)\big(1 - g(F)\big)}{n} \leq \frac{1}{4n}.$$

Seja $\mathcal{P}$ a classe de todas as distribuições absolutamente contínuas na reta real. Sejam $F, G \in \mathcal{P}$ e definamos a função distância $\Delta(F, G)$ como segue:

$$\Delta(F,G) = \int_{-\infty}^{\infty} \big(F(x) - G(x)\big)^2 \frac{F'(x) + G'(x)}{2} \mathrm{d}x \cdot$$

Esta função satisfaz as seguintes propriedades:

- $\Delta(F, G) = 0$ se, e somente se, $F = G$.
- $\Delta(F, G) = \Delta(G, F)$.
- $\Delta(F, G) > 0$.

Exemplo:

Encontremos um estimador não viciado de mínima variância para $\Delta(F,G)$. Sejam $X_1, X_2, \cdots, X_m$ uma amostra aleatória de F e $Y_1, Y_2, \cdots, Y_n$ uma amostra aleatória de G, independentes. Consideramos que $F, G \in \mathcal{P}$. Primeiro mostramos que

$$\begin{aligned} g(F,G) &= P\Big(\{\max(X_1,X_2) < \min(Y_1,Y_2)\}\bigcup \\ &\qquad \{\max(Y_1,Y_2) < \min(X_1,X_2)\}\Big) = \frac{1}{3} + 2\Delta(F,G)\cdot \end{aligned}$$

Temos que

$$\begin{aligned} g(F,G) &= P\Big(\{\max(X_1,X_2) < \min(Y_1,Y_2)\}\Big) \\ &\quad + P\Big(\{\max(Y_1,Y_2) < \min(X_1,X_2)\}\Big)\cdot \end{aligned}$$

Exemplo (continuação):

Para utilizarmos os teoremas acima, vamos definir

$$\varphi(X_1, X_2, Y_1, Y_2) = \begin{cases} 1, & \text{se } \max(X_1, X_2) < \min(Y_1, Y_2) \text{ ou se} \\ & \quad \max(Y_1, Y_2) < \min(X_1, X_2) \\ 0, & \text{caso contrário} \end{cases}$$

Então $\varphi(X_1, X_2, Y_1, Y_2)$ é um estimador não viciado de $g(F, G)$ e de fato é um kernel de $g(F, G)$.

Exemplo (continuação):

A U-estatística correspondente, portanto, deve ser o estimador não viciado de mínima variância. Nós temos

$$U(X,Y) = \frac{1}{\binom{m}{2}\binom{n}{2}} \sum_{i_1<i_2} \sum_{k_1<k_2} \varphi(X_{i_1}, X_{i_2}, Y_{k_1}, Y_{k_2}),$$

de maneira que U é o estimador não viciado de mínima variância de $g(F,G)$ assim como o estimador não viciado de mínima variância de $\Delta(F,G)$ é

$$\widehat{\Delta}(F,G) = \frac{1}{2}U(X,Y) - \frac{1}{6}.$$

Exemplo:

Seja $\mathcal{P}$ a classe de todas as funções de distribuição absolutamente contínuas na reta real e $X_1, X_2, \cdots, X_m$ e $Y_1, Y_2, \cdots, Y_n$ duas amostras aleatórias independentes de F e G, respectivamente, com $F, G \in \mathcal{P}$. Queremos estimar

$$\rho(F, G) = P(X < Y)\cdot$$

Com esse objetivo, vamos definir

$$Z_{ij} = \begin{cases} 1, & X_i < Y_j \\ 0, & X_i \geq Y_j \end{cases}$$

para cada par X_i, Y_j, $i = 1, 2, \cdots, m$ e $j = 1, 2, \cdots, n$.

Exemplo (continuação):

Então $\sum_{i=1}^{m} Z_{ij}$ é o número de vezes que $X < Y_j$ e $\sum_{j=1}^{n} Z_{ij}$ é o número de vezes que $X_i < Y$.

Mann and Whitney (1947) sugeriram utilizar o estimador $\frac{U}{mn}$, onde

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} Z_{ij}$$

e

$$\mathrm{E}(U) = mn\,\mathrm{E}(Z_{ij}) = mn\,P(X < Y)\cdot$$

Então

$$\widehat{\rho}(F, G) = \frac{U}{mn},$$

é não viciado para ρ.