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

Suponha que $f(x;\theta)$, $\theta \in \Omega$, seja uma função densidade ou função de probabilidade. Considere $Y_n = u(X_1, \ldots, X_n)$ baseado em uma amostra aleatória $X_1, \ldots, X_n$, com densidade ou função de probabilidade $f(x;\theta)$.

Suponha que $f(x;\theta)$, $\theta \in \Omega$, seja uma função densidade ou função de probabilidade. Considere $Y_n = u(X_1, \ldots, X_n)$ baseado em uma amostra aleatória $X_1, \ldots, X_n$, com densidade ou função de probabilidade $f(x;\theta)$.

Definição:

Para um dado inteiro positivo n, $Y = u(X_1, X_2, \ldots, X_n)$ é um estimador não viesado de variância uniformemente mínima(ENVVUM) para θ se Y for não viesado, ou seja, $\mathbb{E}(Y) = \theta$, e se a variância de Y for menor ou igual à variância de qualquer outro estimador não viesado de θ.

Observação Importante:

Vamos agora discutir o problema da estimação pontual de um parâmetro a partir de uma perspectiva ligeiramente diferente.

Função de Decisão

Seja $Y = u(X_1, X_2, \ldots, X_n)$ uma estatística com valor observado $y = u(x_1, x_2, \ldots, x_n)$. Seja $\delta(y)$ uma função da estatística observada y, uma estimativa pontual para θ. Assim, a função $\delta(y)$ decide o valor de θ. δ é chamada de função de decisão ou regra de decisão.

Função de Decisão

Seja $Y = u(X_1, X_2, \ldots, X_n)$ uma estatística com valor observado $y = u(x_1, x_2, \ldots, x_n)$. Seja $\delta(y)$ uma função da estatística observada y, uma estimativa pontual para θ. Assim, a função $\delta(y)$ decide o valor de θ. δ é chamada de função de decisão ou regra de decisão.

Um valor da função de decisão, digamos $\delta(y)$, é chamado de decisão. Assim, uma estimativa pontual numericamente determinada de um parâmetro θ é uma decisão. Uma decisão pode estar correta ou pode estar errada. Seria útil ter uma medida da diferença, se houver, entre o valor verdadeiro de θ e a estimativa pontual $\delta(y)$.

Associamos a cada par $(\theta, \delta(y))$, $\theta \in \Omega$, um número não negativo $L(\theta, \delta(y))$ que reflete o quanto $\delta(y)$ está afastado de θ.

- Chamamos a função L de função perda.

Associamos a cada par $(\theta, \delta(y))$, $\theta \in \Omega$, um número não negativo $L(\theta, \delta(y))$ que reflete o quanto $\delta(y)$ está afastado de θ.

- Chamamos a função L de função perda.

Para a variável aleatória contínua Y, podemos calcular a função Risco:

$$R(\theta, \delta) = \mathbb{E}\Big(L(\theta, \delta(y))\Big),$$

em que,

- $\mathbb{E}\Big(L(\theta, \delta(y))\Big) = \int_{-\infty}^{\infty} L(\theta, \delta(y)) \cdot f_Y(y) dy$

Exemplo

$X_1, X_2, \ldots, X_{25} \overset{\text{iid}}{\sim} N(\theta, 1)$, $\theta \in \mathbb{R}$. Seja $Y = \bar{X} \overset{\text{iid}}{\sim} N(\theta, \dfrac{1}{25})$. Considere, $\delta_1(y) = y$, $\delta_2(y) = 0$ e tome $L(\theta, \delta(y)) = (\theta - \delta(y))^2$. As funções de risco correspondentes são:

- $R(\theta, \delta_1) = E[(\theta - Y)^2] = Var(Y) = \frac{1}{25}$;
- $R(\theta, \delta_2) = E[(\theta - 0)^2] = \theta^2$.

Exemplo

$X_1, X_2, \ldots, X_{25} \overset{\text{iid}}{\sim} N(\theta, 1)$, $\theta \in \mathbb{R}$. Seja $Y = \bar{X} \overset{\text{iid}}{\sim} N(\theta, \frac{1}{25})$. Considere, $\delta_1(y) = y$, $\delta_2(y) = 0$ e tome $L(\theta, \delta(y)) = (\theta - \delta(y))^2$. As funções de risco correspondentes são:

- $R(\theta, \delta_1) = E[(\theta - Y)^2] = Var(Y) = \frac{1}{25}$;
- $R(\theta, \delta_2) = E[(\theta - 0)^2] = \theta^2$.

Se nosso critério for selecionar o estimador com menor risco, temos:
$R(\theta, \delta_2) \leq R(\theta, \delta_1) \iff \theta^2 \leq \dfrac{1}{25} \iff -\dfrac{1}{5} \leq \theta \leq \dfrac{1}{5}$.

Exemplo

$X_1, X_2, \ldots, X_{25} \stackrel{\text{iid}}{\sim} N(\theta, 1)$, $\theta \in \mathbb{R}$. Seja $Y = \bar{X} \stackrel{\text{iid}}{\sim} N(\theta, \frac{1}{25})$. Considere, $\delta_1(y) = y$, $\delta_2(y) = 0$ e tome $L(\theta, \delta(y)) = (\theta - \delta(y))^2$. As funções de risco correspondentes são:

- $R(\theta, \delta_1) = E[(\theta - Y)^2] = Var(Y) = \frac{1}{25}$;
- $R(\theta, \delta_2) = E[(\theta - 0)^2] = \theta^2$.

Se nosso critério for selecionar o estimador com menor risco, temos:
$R(\theta, \delta_2) \leq R(\theta, \delta_1) \iff \theta^2 \leq \dfrac{1}{25} \iff -\dfrac{1}{5} \leq \theta \leq \dfrac{1}{5}$.

Neste caso, se $-\dfrac{1}{5} \leq \theta \leq \dfrac{1}{5}$ o estimador δ_2 será melhor que o estimador δ_1. Caso contrário, ou seja, para $\theta \notin [-\dfrac{1}{5}, \dfrac{1}{5}]$, δ_1 será melhor que δ_2.

Agora, vamos nos restringir a classe de estimadores não viesados ($\mathbb{E}[\delta(Y)] = \theta$). Utilizando a perda quadrática $L(\theta, \delta(y)) = (\theta - \delta(y)^2)$, escolher um estimador não viesado para θ que tem o menor risco dentre todos os estimadores não viesados para θ. Nos deparamos com o problema de encontrar o ENVVUM (estimador não tendencioso de variância uniformemente mínima).

Poderíamos também adotar outro critério. Em vez de nos restringir a classe dos estimadores não viesados, poderíamos escolher o estimador que minimiza o máximo risco ($\max_\theta R(\theta, \delta)$), conhecido como estimador Minimax. Com esse critério, nosso exemplo, $\max_\theta R(\theta, \delta_1) = \max_\theta \left(\frac{1}{25}\right) = \frac{1}{25}$ e $\max_\theta R(\theta, \delta_2) = +\infty$ e, portanto, o melhor estimador seria o estimador δ_1.

Poderíamos também adotar outro critério. Em vez de nos restringir a classe dos estimadores não viesados, poderíamos escolher o estimador que minimiza o máximo risco ($\max_\theta R(\theta, \delta)$), conhecido como estimador Minimax. Com esse critério, nosso exemplo, $\max_\theta R(\theta, \delta_1) = \max_\theta \left(\frac{1}{25}\right) = \frac{1}{25}$ e $\max_\theta R(\theta, \delta_2) = +\infty$ e, portanto, o melhor estimador seria o estimador δ_1.

As funções perda mais utilizadas:

- $L(\theta, \delta(y)) = (\theta - \delta(y))^2$, conhecida como perda quadrática.
- $L(\theta, \delta(y)) = |\theta - \delta(y)|$, conhecida como erro absoluto.

Neste exemplo, ilustramos o seguinte:

1. Sem alguma restrição sob a função de decisão, é difícil encontrar uma função de decisão que tem risco uniformemente menor que outras funções de decisão.

Neste exemplo, ilustramos o seguinte:

1. Sem alguma restrição sob a função de decisão, é difícil encontrar uma função de decisão que tem risco uniformemente menor que outras funções de decisão.
2. Um princípio de seleção da melhor função de decisão é chamado de **princípio minimax**. Esse princípio diz que:

Se a função de decisão dada por $\delta_0(y)$ é tal que, para todo $\theta \in \Omega$, $\max_\theta R[\theta, \delta_0(y)] \leq \max_\theta R[\theta, \delta(y)]$ para qualquer outra função de decisão $\delta(y)$, então $\delta_0(y)$ é chamada de função de decisão minimax.

Exercícios

Para 🏠

Exercícios 7.1: 1,2,4 ao 9

Estatística Suficiente para um Parâmetro

Definição 1

Sejam $X_1, X_2, \ldots, X_n$ uma amostra aleatória com densidade ou função de probabilidade $f(x;\theta)$, $\theta \in \Omega$. Seja $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ uma estatística cuja função de densidade ou função de probabilidade é $f_{Y_1}(y_1;\theta)$. Então, Y_1 é uma estatística suficiente para θ se, e somente se,

$$\frac{f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)}{f_{Y_1}(y_1;\theta)} = H(x_1, x_2, \ldots, x_n), \tag{1}$$

em que $H(x_1, x_2, \ldots, x_n)$ não depende de $\theta \in \Omega$.

Exemplo

$X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \Gamma(2, \theta),\ \theta > 0$. Seja $y = \sum_{i=1}^{n} X_i$. Mostre que $Y \overset{\text{iid}}{\sim} \Gamma(2n, \theta)$ usando função geradora de momentos.

Exemplo

$X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \Gamma(2, \theta),\ \ \theta > 0$. Seja $y = \sum_{i=1}^{n} X_i$. Mostre que $Y \overset{\text{iid}}{\sim} \Gamma(2n, \theta)$ usando função geradora de momentos.

Como a função geradora de momentos associada a essa distribuição é dada por $M(t) = (1 - \theta t)^{-2}, t < \frac{1}{\theta}$, a função geradora de momentos de $Y = X_1 + X_2 + \ldots + X_n$ é:

$$
\begin{aligned}
E[e^{t(X_1+X_2+\ldots+X_n)}] &= E[e^{tX_1} e^{tX_2} \ldots e^{tX_n}] \\
&= [(1 - \theta t)^{-2}]^n \\
&= (1 - \theta t)^{-2n}.
\end{aligned}
$$

Logo, $Y \overset{\text{iid}}{\sim} \Gamma(2n, \theta)$.

Notem que, as funções de densidade $f_{X_1}(x_1;\theta), f_Y(y;\theta)$ são definidas como:

$$f_{X_1}(x_1;\theta) = \frac{1}{\Gamma(2)\theta^2} x_1^{2-1} e^{-\frac{x}{\theta}},\ x_1 > 0.$$

$$f_Y(y;\theta) = \frac{1}{\Gamma(2n)\theta^{2n}} y^{2n-1} e^{-\frac{y}{\theta}},\ y = \sum_{i=1}^{n} X_i.$$

Segue que,

$$\frac{\frac{x_1^{2-1}e^{-x_1/\theta}}{\Gamma(2)\theta^2}\cdot\frac{x_2^{2-1}e^{-x_2/\theta}}{\Gamma(2)\theta^2}\cdot\ldots\cdot\frac{x_n^{2-1}e^{-x_n/\theta}}{\Gamma(2)\theta^2}}{\frac{(x_1+x_2+\ldots+x_n)^{2n-1}e^{-(x_1+x_2+\ldots+x_n)/\theta}}{\Gamma(2n)\theta^{2n}}} = \frac{\Gamma(2n)}{(\Gamma(2))^n}\cdot\frac{x_1x_2\cdot\ldots\cdot x_n}{(x_1+x_2+\ldots+x_n)^{2n-1}}$$

Como, a expressão final, não depende de θ, Y é uma estatística suficiente para θ.

Teorema de Neyman (Critério de Fatoração)

Teorema 1

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória com densidade de probabilidade ou função de probabilidade $f(x;\theta)$, $\theta \in \Omega$. A estatística $Y_1 = u(X_1, \ldots, X_n)$ é uma estatística suficiente para θ se, e somente se, existir duas funções não negativas, k_1 e k_2, tais que

$$f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta) = k_1[u_1(x_1, x_2, \ldots, x_n);\theta]k_2(x_1, x_2, \ldots, x_n)$$

em que $k_2(x_1, x_2, \ldots, x_n)$ não depende de θ.

Observação (Jacobiano):

Seja (X_1, X_2) um par aleatório absolutamente contínuo com densidade de probabilidade conjunta $f_{X_1,X_2}(x_1, x_2)$. Seja também $U : \mathbb{R}^2 \to \mathbb{R}^2$ uma função injetiva (portanto com inversa) com dois componentes $U(x_1, x_2) = (y_1, y_2)$. Cada um destes componentes é função de duas variáveis reais, tal que:

$$\begin{cases} y_1 = u_1(x_1, x_2) \\ y_2 = u_2(x_1, x_2) \end{cases}$$

sendo que u_1 e u_2 possuem derivadas parciais em relação a x_1 e x_2. Portanto, podemos definir o par aleatório $(Y_1, Y_2) = U(X_1, X_2)$. Como determinar a densidade de probabilidade conjunta do par (Y_1, Y_2) a partir da densidade conjunta de (X_1, X_2)?

Observação (Jacobiano):

Como U tem inversa, podemos escrever:

$$\begin{cases} x_1 = w_1(y_1, y_2) \\ x_2 = w_2(y_1, y_2) \end{cases}$$

A densidade conjunta de (Y_1, Y_2) será:

$$f_{Y_1,Y_2}(y_1, y_2) = |J| \, f_{X_1,X_2}\left(w_1(y_1, y_2), w_2(y_1, y_2)\right)$$

em que $|J|$ representa o módulo do determinante jacobiano, isto é, o módulo de:

$$\left| \begin{vmatrix} \frac{\partial w_1(y_1,y_2)}{\partial y_1} & \frac{\partial w_1(y_1,y_2)}{\partial y_2} \\ \frac{\partial w_2(y_1,y_2)}{\partial y_1} & \frac{\partial w_2(y_1,y_2)}{\partial y_2} \end{vmatrix} \right| .$$

Demonstração

Considere o caso contínuo!

Faremos a "volta" do teorema primeiro ($\Leftarrow$). Por hipótese, existem duas funções não negativas, k_1 e k_2, tais que

$$f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta) = k_1[u_1(x_1,x_2,\dots,x_n);\theta]k_2(x_1,x_2,\dots,x_n),$$

ou seja,

$$\frac{f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta)}{f_{Y_1}(y_1,\theta)} = \frac{k_1[u_1(x_1,x_2,\dots,x_n);\theta]k_2(x_1,x_2,\dots,x_n)}{f_{Y_1}(y_1,\theta)}$$

Demonstração

Considere o caso contínuo!

Faremos a "volta" do teorema primeiro ($\Leftarrow$). Por hipótese, existem duas funções não negativas, k_1 e k_2, tais que

$$f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta) = k_1[u_1(x_1,x_2,\dots,x_n);\theta]k_2(x_1,x_2,\dots,x_n),$$

ou seja,

$$\frac{f(x_1;\theta)f(x_2;\theta)\dots f(x_n;\theta)}{f_{Y_1}(y_1,\theta)} = \frac{k_1[u_1(x_1,x_2,\dots,x_n);\theta]k_2(x_1,x_2,\dots,x_n)}{f_{Y_1}(y_1,\theta)}$$

Precisamos encontrar $f_{Y_1}(y_1,\theta)$.

Considere o caso contínuo!

Considere

$$y_1 = u_1(x_1, x_2, \ldots, x_n), \ldots, y_n = u_n(x_1, x_2, \ldots, x_n)$$

tendo as funções inversas,

$$x_1 = w_1(y_1, y_2, \ldots, y_n), \ldots, x_n = w_n(y_1, y_2, \ldots, y_n)$$

com Jacobiano J; Sabemos que,

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_n}(x_n).$$

Para simplificar a notação considere $w_i = w_i(y_1, y_2, \ldots, y_n), \forall i = 1, 2, \cdots, n$.

Com isso,

$$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = f_{X_1}(w_1)f_{X_2}(w_2)\ldots f_{X_n}(w_n)|J|.$$

Logo, por hipótese, existem duas funções não negativas, k_1 e k_2, tais que

$$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = |J|k_1[y_1;\theta]k_2(w_1, w_2, \ldots, w_n)$$

Daí,

$$\begin{aligned} f_{Y_1}(y_1) &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} k_1[y_1;\theta]k_2(w_1, w_2, \ldots, w_n)|J|dy_2dy_3\cdots dy_n \\ &= k_1(y_1,\theta)\underbrace{\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} k_2(w_1, w_2, \ldots, w_n)|J|dy_2dy_3\cdots dy_n}_{\text{Não depende de } \theta, \text{ somente de } y_1 (=m(y_1))} \end{aligned}$$

Com isso,

$$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = f_{X_1}(w_1) f_{X_2}(w_2) \ldots f_{X_n}(w_n)|J|.$$

Logo, por hipótese, existem duas funções não negativas, k_1 e k_2, tais que

$$f_{Y_1,Y_2,\ldots,Y_n}(y_1, y_2, \ldots, y_n) = |J| k_1[y_1;\theta] k_2(w_1, w_2, \ldots, w_n)$$

Daí,

$$\begin{aligned} f_{Y_1}(y_1) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} k_1[y_1;\theta] k_2(w_1, w_2, \ldots, w_n)|J| dy_2 dy_3 \cdots dy_n \\ &= k_1(y_1,\theta) \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} k_2(w_1, w_2, \ldots, w_n)|J| dy_2 dy_3 \cdots dy_n}_{\text{Não depende de } \theta \text{, somente de } y_1 (=m(y_1))} \end{aligned}$$

Então, $f_{Y_1}(y_1) = k_1(y_1,\theta) m(y_1)$

Com isso,

$$\frac{f(x_1;\theta)\cdot f(x_2;\theta)\dots f(x_n;\theta)}{f_{Y_1}(y_1,\theta)} = \frac{\cancel{k_1(y_1,\theta)}k_2(x_1,x_2,\dots,x_n)}{\cancel{k_1(y_1,\theta)}m(y_1)}$$

que não depende de θ, portanto, Y é suficiente para θ. ■

$(\Rightarrow)$

Por suposição,

$$f(x_1;\theta)\cdot f(x_2;\theta)\dots f(x_n;\theta) = \underbrace{H(x_1,\dots,x_n)}_{k_2}\underbrace{f_{Y_1}(y_1,\theta)}_{k_1}. \blacksquare$$

Exemplo

Seja $X_1, X_2, \ldots, X_n$ amostra aleatória com densidade $f(x;\theta) = \theta x^{\theta-1}$, $x \in (0,1)$ e $\theta > 0$.

$$
\begin{aligned}
f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta) &= \theta^n x_1^{\theta-1}\ldots x_n^{\theta-1} \\
&= \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta} \frac{1}{\displaystyle\prod_{i=1}^{n} x_i} \\
&= k_1(\theta, \prod_{i=1}^{n} x_i) k_2(x_1, \ldots, x_n)
\end{aligned}
$$

Uma vez que $k_2(x_1, x_2, \ldots, x_n)$ não depende de θ, o produto $\prod_{i=1}^{n} X_i$, pelo critério da fatoração, é uma estatística suficiente para θ.

Vantagens de uma Estatística Suficiente

Saber que uma estatística é suficiente oferece as seguintes vantagens:

- Reduz a quantidade de dados que precisamos manipular, sem perda de informação.

Vantagens de uma Estatística Suficiente

Saber que uma estatística é suficiente oferece as seguintes vantagens:

- Reduz a quantidade de dados que precisamos manipular, sem perda de informação.
- Melhora a eficiência dos estimadores, ajudando a encontrar estimadores de menor variância.

Vantagens de uma Estatística Suficiente

Saber que uma estatística é suficiente oferece as seguintes vantagens:

- Reduz a quantidade de dados que precisamos manipular, sem perda de informação.
- Melhora a eficiência dos estimadores, ajudando a encontrar estimadores de menor variância.
- Facilita a fatoração da função de verossimilhança, simplificando cálculos.

Vantagens de uma Estatística Suficiente

Saber que uma estatística é suficiente oferece as seguintes vantagens:

- Reduz a quantidade de dados que precisamos manipular, sem perda de informação.
- Melhora a eficiência dos estimadores, ajudando a encontrar estimadores de menor variância.
- Facilita a fatoração da função de verossimilhança, simplificando cálculos.
- Auxilia na construção de testes e intervalos de confiança ótimos.

Vantagens de uma Estatística Suficiente

Saber que uma estatística é suficiente oferece as seguintes vantagens:

- Reduz a quantidade de dados que precisamos manipular, sem perda de informação.
- Melhora a eficiência dos estimadores, ajudando a encontrar estimadores de menor variância.
- Facilita a fatoração da função de verossimilhança, simplificando cálculos.
- Auxilia na construção de testes e intervalos de confiança ótimos.
- Garante a melhor inferência possível em termos de uso da informação dos dados.

Para 🏠

Exercícios 7.2: 1,2,4 ao 8

Referências I

- Bolfarine, Heleno e Mônica Carneiro Sandoval (2001). *Introdução à inferência estatística*. Vol. 2. SBM.
- Casella, George e Roger L Berger (2021). *Statistical inference*. Cengage Learning.
- Hogg, RV, J McKean e AT Craig (2019). *Introduction to Mathematical Statistics*.