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

Caso Multiparamétrico: Estimação

Considere θ um vetor de p parâmetros. Sejam $X_1, \ldots, X_n$ variáveis aleatórias iid com densidade $f(x, \theta)$, $\theta \in \Omega \subset \mathbb{R}^{\mathrm{p}}$. A função log-verossimilhança é dada por

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i, \theta)$$

A teoria para o caso multiparamétrico requer condições de regularidade adicionais. Considere o apêndice do livro texto Hogg et al. (2019), página 687.

Observação:

A prova do teorema abaixo não depende se θ é um escalar ou um vetor. Portanto, o resultado é válido também para o caso multiparamétrico.

Teorema 1

Seja θ_0 o valor verdadeiro de θ. Sob $(R0)$ e $(R1)$, temos que

$$\lim_{n\to\infty} P_{\theta_0}(L(\theta_0, X) > L(\theta, X)) = 1, \ \textit{para } \theta \neq \theta_0,$$

ou seja, θ_0 é o ponto de máximo de $L(\theta, X)$.

Com isso, usaremos como estimador de θ aquele valor que maximiza $L(\theta)$, ou equivalentemente, o valor que resolve o sistema de equações $\dfrac{\partial \ell(\theta)}{\partial \theta} = 0$. Se esse valor existir, será chamado de estimador de máxima verossimilhança (EMV).

Com isso, usaremos como estimador de θ aquele valor que maximiza $L(\theta)$, ou equivalentemente, o valor que resolve o sistema de equações $\dfrac{\partial \ell(\theta)}{\partial \theta} = 0$. Se esse valor existir, será chamado de estimador de máxima verossimilhança (EMV).

Se o interesse for estimar uma função de θ, digamos $\eta = g(\theta)$ e se $\hat{\theta}$ é um EMV para θ, então $\hat{\eta} = g(\hat{\theta})$ é um estimador de máxima verossimilhança para η.

Exemplo

$X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$. Neste caso, $\theta = (\mu, \sigma)$.

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}, \theta \in \Omega = \mathbb{R} \times (0, +\infty) \text{ e } \mathrm{x} \in \mathbb{R}.$$

Exemplo

$X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$. Neste caso, $\theta = (\mu, \sigma)$.

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \theta \in \Omega = \mathbb{R} \times (0, +\infty) \text{ e } x \in \mathbb{R}.$$

A função de verossimilhança nesse caso é:

$$L(\theta) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}$$

$$\ell(\theta) = \log L(\theta) = -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

fazendo $\frac{\partial \ell(\theta)}{\partial \mu} = 0$ e $\frac{\partial \ell(\theta)}{\partial \sigma} = 0$, pode - se obter:

$$\hat{\mu} = \bar{X} \text{ e } \hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

Analisando a matriz de segundas derivadas de ℓ, pode-se ver que $(\hat{\mu}, \hat{\sigma})$ é o ponto de máximo global de $\ell(\mu, \sigma)$. Ou seja, $(\hat{\mu}, \hat{\sigma})$ é o EMV de (μ, σ).

No caso univariado, a informação de Fisher é a variância da v.a. $\dfrac{\partial \log f(x,\theta)}{\partial \theta}$. No caso multiparamétrico, a informação de Fisher é definida como a matriz de variância e covariância do vetor aleatório

$$\bigtriangledown \log f(\boldsymbol{X},\theta) = \Big(\frac{\partial \log f(x,\theta)}{\partial \theta_1},\ldots,\frac{\partial \log f(x,\theta)}{\partial \theta_n}\Big)^{\top}$$

A informação de Fisher é $I(\theta) = Cov(\bigtriangledown \log f(\boldsymbol{X},\theta))$, em que

$$Cov(Z_1,Z_2) = \begin{bmatrix} Cov(Z_1,Z_1) & Cov(Z_1,Z_2) \\ Cov(Z_2,Z_1) & Cov(Z_2,Z_2) \end{bmatrix} = \begin{bmatrix} Var(Z_1) & Cov(Z_1,Z_2) \\ Cov(Z_2,Z_1) & Var(Z_2) \end{bmatrix}$$

O elemento (j, k) de $I(\theta)$ é dado por

$$I_{jk}(\theta) = Cov\Big(\frac{\partial \log f(x,\theta)}{\partial \theta_j}, \frac{\partial \log f(x,\theta)}{\partial \theta_k}\Big)$$

O elemento (j, k) de $I(\theta)$ é dado por

$$I_{jk}(\theta) = Cov\Big(\frac{\partial \log f(x, \theta)}{\partial \theta_j}, \frac{\partial \log f(x, \theta)}{\partial \theta_k}\Big)$$

Sob as condições de regularidade, temos que

$$1 = \int_{-\infty}^{+\infty} f(x, \theta) dx \Rightarrow 0 = \frac{\partial}{\partial \theta} \int_{-\infty}^{+\infty} f(x, \theta) dx$$
$$\Rightarrow 0 = \int_{-\infty}^{+\infty} \frac{\partial f(x, \theta)}{\partial \theta} dx = \int_{-\infty}^{+\infty} \frac{\partial \log f(x, \theta)}{\partial \theta} f(x, \theta) dx$$
$$= E\Big[\frac{\partial \log f(x, \theta)}{\partial \theta}\Big] = E\Big[\bigtriangledown \log f(x, \theta)\Big]$$

Tomando mais uma derivada com relação a θ, obtemos:

$$0_{1\times p}^{\top} = \frac{\partial}{\partial\theta^{\top}}\int_{-\infty}^{+\infty}\frac{\partial \log f(x,\theta)}{\partial\theta} f(x,\theta)dx$$

$$= \int \frac{\partial^2 \log f(x,\theta)}{\partial\theta\partial\theta^{\top}} f(x,\theta)dx + \int \frac{\partial \log f(x,\theta)}{\partial\theta}\frac{\partial \log f(x,\theta)}{\partial\theta^{\top}} f(x,\theta)dx$$

$$\Rightarrow -\int \frac{\partial^2 \log f(x,\theta)}{\partial\theta\partial\theta^{\top}} f(x,\theta)dx = \int \frac{\partial \log f(x,\theta)}{\partial\theta}\frac{\partial \log f(x,\theta)}{\partial\theta^{\top}} f(x,\theta)dx$$

$$\Rightarrow \quad E\left(\frac{\partial \log f(x,\theta)}{\partial\theta}\frac{\partial \log f(x,\theta)}{\partial\theta^{\top}}\right) = -E\left(\frac{\partial^2 \log f(x,\theta)}{\partial\theta\partial\theta^{\top}}\right)$$

Daí,

$$I(\theta) = -E\left(\frac{\partial^2 \log f(x,\theta)}{\partial\theta\partial\theta^\top}\right).$$

Exemplo

$X \sim N(\mu, \sigma), \;\; \theta = (\mu, \sigma).$

$$\frac{\partial \log f}{\partial \mu} = \frac{1}{\sigma^2}(x-\mu); \;\; \frac{\partial^2 \log f}{\partial \mu^2} = -\frac{1}{\sigma^2}; \;\; \frac{\partial \log f}{\partial \sigma} = -\frac{1}{\sigma} + \frac{1}{\sigma^3}(x-\mu)^2$$

$$\frac{\partial^2 \log f}{\partial \sigma^2} = \frac{1}{\sigma^2} - \frac{3}{\sigma^4}(x-\mu)^2; \;\; \frac{\partial^2 \log f}{\partial \mu \partial \sigma} = -\frac{2}{\sigma^3}(x-\mu)$$

Logo,

$$I(\theta) = I(\mu, \sigma) = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix} = \begin{bmatrix} -E\left(\frac{\partial^2 \log f(x,\theta)}{\partial \mu^2}\right) & -E\left(\frac{\partial^2 \log f(x,\theta)}{\partial \sigma \partial \mu}\right) \\ E\left(\frac{\partial^2 \log f(x,\theta)}{\partial \mu \partial \sigma}\right) & -E\left(\frac{\partial^2 \log f(x,\theta)}{\partial \sigma^2}\right) \end{bmatrix}$$

Teorema 2

Sejam $X_1, \ldots, X_n$ variáveis aleatórias iid com densidade $f(x, \theta), \theta \in \Omega \subset \mathbb{R}^{\mathrm{P}}$. Assuma que valem as condições de regularidade. Então,

1. *O sistema de equações* $\dfrac{\partial \ell(\theta)}{\partial \theta} = 0$ *tem uma solução* $\hat{\theta}_n$ *tal que* $\hat{\theta}_n \xrightarrow{P} \theta_0$ *(θ_0 verdadeiro valor de θ).*
2. *Para qualquer sequência satisfazendo 1, temos*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N_p(0, I^{-1}(\theta)).$$

Corolário:

Sob as condições do teorema anterior, temos que

$$\sqrt{n}(\hat{\theta}_{in} - \theta_{i0}) \xrightarrow{D} N_p(0, I_{ii}^{-1}(\theta)).$$

Seja $g(\theta) = (g_1(\theta), \ldots, g_k(\theta))^\top, 1 \leq k \leq p$ tal que a matriz de derivadas parciais $B = \left(\frac{\partial g_i}{\partial \theta_j}\right)_{ij}, i = 1, \ldots, k$ e $j = 1, \ldots, p$ possuam elementos contínuos e que não se anulem numa vizinhança de θ_0. Seja $\hat{\eta} = g(\hat{\theta})$. Então, $\hat{\eta}$ é o EMV de η e

$$\sqrt{n}(\hat{\eta} - \eta_0) \xrightarrow{D} N_k(0, BI^{-1}(\theta)B^\top).$$

Para 🏠

- **Exercícios da seção 6.4:** Todos, exceto o 5 e o 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*.