Métodos de Estimação

Professor: Pedro M. Almeida-Junior

24 de fevereiro de 2022

Departamento de Estatística (UEPB)

Métodos de Estimação

- É possível obter muitos estimadores para o mesmo parâmetro a partir de regras de estimação diferentes. Assim, serão abordados a seguir alguns métodos de estimação.

    1. Método da Máxima Verossimilhança

    2. Método dos Momentos

Método da máxima verossimilhança

- Seja $X_1, \ldots, X_n$ uma AAS de uma variável aleatória com função de densidade ou de probabilidade $f(x \mid \theta)$. A função $L(\theta)$, que é

$$L(\theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

que determina, para cada conjunto de valores observados de $X_1, \ldots, X_n$ a probabilidade de se obter estes valores é chamada função de verossimilhança de θ.

- A função de verossimilhança correspondente a uma amostra aleatória observada é definida como sendo igual a função densidade (ou de probabilidade) conjunta, embora seja interpretada diferentemente como função de θ para $X_1, \ldots, X_n$ conhecidos.

- Neste caso, um estimador razoável para θ é aquele que maximiza a chance relativa de se obter o que realmente foi observado. Sendo assim, o valor $\hat{\theta}$ e θ, correspondente ao máximo global de $L(\theta)$, se existir é chamado estimador de máxima verossimilhança (EMV) de θ.

- O logaritmo natural da função de verossimilhança de θ é denotado por

$$l(\theta) = \log L(\theta)$$

OBS: O valor que maximiza $L(\theta)$, também maximiza $l(\theta)$, uma vez que o logaritmo é função decrescente.

No caso uniparamétrico onde Θ é um intervalo da reta e $l(\theta; \mathrm{x})$ é derivável, o estimador de máxima verossimilhança pode ser encontrado como a raiz da equação de verossimilhança

$$l'(\theta; \mathrm{x}) = \left.\frac{\partial l(\theta; \mathrm{x})}{\partial \theta}\right|_{\theta=\hat{\theta}} = 0$$

Em alguns exemplos simples, a solução da equação de verossimilhança pode ser **obtida explicitamente**. Em situações mais complicadas, a solução da equação acima será em geral obtida por **procedimentos numéricos**. Para se concluir que a solução da equação anterior é um ponto de máximo, é necessário verificar se

$$l''(\hat{\theta}; \mathrm{x}) = \left.\frac{\partial^2 \log L(\theta; \mathrm{x})}{\partial \theta^2}\right|_{\theta=\hat{\theta}} < 0$$

o estimador de Máxima Verossimilhança é função de uma estatística suficiente

O teorema a seguir estabelece que o estimador de Máxima Verossimilhança é função de uma estatística suficiente.

Teorema: Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória X com função de densidade (ou de probabilidade) $f(x \mid \theta)$. Seja $T = T(X_1, \ldots, X_n)$ uma estatística suficiente para θ. Então o estimador de máxima verossimilhança $\hat{\theta}$ (se existir) é função de T.

Princípio da Invariância

Teorema (O princípio da invariância): Seja $g(.)$ uma função real 1:1 (inversível) definida em Θ. Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória X com função de densidade (ou de probabilidade) $f(x \mid \theta)$. Se $\hat{\theta}$ é um estimador de máxima verossimilhança de θ, então $g(\hat{\theta})$ é um estimador de máxima verossimilhança de $g(\theta)$.

Exemplos

Exemplo 1: Suponha $X_1, \ldots, X_n$ uma AAS de uma v.a. $X \sim N(\mu, \sigma^2)$, com σ^2 conhecido. Encontre o EMV de μ.

Exemplo 2: Sejam $X_1, \ldots, X_n$ a.a. da v.a. $X \sim N(\mu, \sigma^2)$. Aplique o princípio da da invariância do EMV.

Suponha agora que temos uma AAS de tamanho n, da v.a. X, ou seja, $X_1, \ldots, X_n$, e que a distribuição de X envolve o parâmetro θ. Suponha ainda que X satisfaz as condições de regularidade exigidas no Teorema de Crámer-Rao, então, se $\hat{\theta}$ é o **estimador de Máxima Verossimilhança** de θ, temos que,

$$\sqrt{n}(\hat{\theta} - \theta) \overset{a}{\sim} N\left(0, \frac{1}{I_F(\theta)}\right),$$

e

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \overset{a}{\sim} N\left(0, \frac{(g'(\theta))^2}{I_F(\theta)}\right)$$

em que "$\overset{a}{\sim}$"significa **distribuição assintótica**.

Propriedades assintóticas do EMV

Para amostras grandes,

- Os estimadores de máxima verossimilhança de θ e $g(\theta)$ são **aproximadamente não viciados**,
- As variâncias coincidem com os correspondentes limites inferiores das variâncias dos estimadores não viciados de θ e $g(\theta)$. Portanto, temos que **o estimador de máxima verossimilhança é eficiente**.

EMV para o Caso multiparamétrico

- É possível generalizar os resultados da estimação por máxima verossimilhança para o caso multiparamétrico.

- Assim, pela distribuição de X envolve p parâmetros, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$, então a função de verossimilhança associada a uma amostra aleatória, $X_1, \ldots, X_n$ é

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i \mid \boldsymbol{\theta})$$

em que $p < n$.

Vetor Escore e Matriz de Informação de Fisher

- A função escore é substituída pelo **vetor escore** da v.a. X.

$$U(\theta_1, \ldots, \theta_p) = \begin{bmatrix} \frac{\partial}{\partial \theta_1} \log f(\boldsymbol{x}, \theta_1, \ldots, \theta_p) \\ \vdots \\ \frac{\partial}{\partial \theta_p} \log f(\boldsymbol{x}, \theta_1, \ldots, \theta_p) \end{bmatrix}$$

- A informação de Fisher é substituída pela **matriz de informação de Fisher**:

$$K(\theta_1, \ldots, \theta_p) = -\mathbb{E} \begin{bmatrix} \frac{\partial^2}{\partial \theta_1^2} \log f(\boldsymbol{x}, \boldsymbol{\theta}) & \cdots & \frac{\partial^2}{\partial \theta_1 \theta_p} \log f(\boldsymbol{x}, \boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial \theta_p \theta_1} \log f(\boldsymbol{x}, \boldsymbol{\theta}) & \cdots & \frac{\partial^2}{\partial \theta_p^2} \log f(\boldsymbol{x}, \boldsymbol{\theta}) \end{bmatrix}$$

Propriedades Assintóticas do EMV caso multiparamétrico

Sob certas condições de regularidade vale o teorema da normalidade assintótica:

$$\sqrt{n}(\hat{\theta} - \theta) \overset{a}{\sim} N_p\left(0, \frac{1}{K(\boldsymbol{\theta})}\right),$$

em que N_p denota a distribuição normal p variada e $\hat{\boldsymbol{\theta}}$ obtido da **maximização da log-verossimilhança**, ou seja, equivalentemente, como solução do sistema

$$U(\theta_1, \ldots, \theta_p) = \mathbf{0}$$

Verossimilhança para amostras independentes

- Suponha que temos **duas ou mais amostras independentes** de distribuições que dependem de um parâmetro, ou vetor de parâmetro.

- Por simplicidade, seja o caso de duas amostras independentes, ficando claro como estender para várias amostras independentes. Assim, sejam $X_1, \ldots, X_n$ e $Y_1, \ldots, Y_n$ duas amostras independentes, então,

$$L(\boldsymbol{\theta}, x_1, \ldots, x_n, y_1, \ldots, y_n) = L(\boldsymbol{\theta}, x_1, \ldots, x_n)L(\boldsymbol{\theta}, y_1, \ldots, y_n)$$

  em que $L(\boldsymbol{\theta}, x_1, \ldots, x_n)$ é a verossimilhança da amostra $X_1, \ldots, X_n$ e $L(\boldsymbol{\theta}, y_1, \ldots, y_n)$ é a verossimilhança da amostra $Y_1, \ldots, Y_n$.

log-verossimilhança para amostras independentes

Daí, a log-verossimilhança é dada por

$$\ell(\boldsymbol{\theta}, x_1, \ldots, x_n, y_1, \ldots, y_n) = \ell(\boldsymbol{\theta}, x_1, \ldots, x_n) + \ell(\boldsymbol{\theta}, y_1, \ldots, y_n)$$

de tal forma que o logaritmo da verossimilhança conjunta é igual a **soma do logaritmo das verossimilhanças** correspondentes a cada uma das amostras.

Exemplo: Sejam $X_1, \ldots, X_n$ uma amostra aleatória correspondente a $X \sim N(\mu, 4)$ e $Y_1, \ldots, Y_n$ uma amostra aleatória correspondente a $Y \sim N(\mu, 9)$. Assumindo que as duas amostras são independentes.

Estimação pelo Método dos Momentos

Método dos momentos

- O método dos momentos é um dos métodos de estimação mais antigos e simples

- Suponha que $X_1, \ldots, X_n$ amostra aleatória de uma distribuição indexada por p parâmetros ($\theta_1, \ldots, \theta_p$, $p < n$). Consideremos os p primeiros momentos da distribuição, $\mathbb{E}(X^k)$, $k = 1, \ldots, n$.

Por exemplo, se a distribuição é contínua com densidade $f(x, \boldsymbol{\theta})$, então,

$$\mathbb{E}(X^k) \int_{-\infty}^{\infty} x^k f(x, \boldsymbol{\theta}) dx,$$

$\mathbb{E}(X), \ldots, \mathbb{E}(X^p)$ são funções de parâmetros desconhecidos

$$\mathbb{E}(X^k) = M_k(\theta_1, \ldots, \theta_p), \forall k = 1, \ldots, p$$

Pela **Lei dos grandes números**, devemos esperar que $\frac{1}{n} \sum_{i=1}^{n} x_i^k$ seja próximo de $\mathbb{E}(X^k) = M_k$ se n é grande.

- O **método dos momentos** consiste em definir o seguinte conjunto de equações:

$$\begin{cases} \frac{1}{n}\sum\limits_{i=1}^{n} x_i = M_1(\theta_1, \ldots, \theta_p) \\ \vdots \\ \frac{1}{n}\sum\limits_{i=1}^{n} x_i^p = M_p(\theta_1, \ldots, \theta_p) \end{cases}$$

e através da solução do sistema acima, encontrar estimadores para os parâmetros $\theta_1, \ldots, \theta_p$, diga-se, $\hat{\theta}_1, \ldots, \hat{\theta}_p$.

Exemplos

Exemplo: Suponha $X_1, \ldots, X_n$ uma amostra aleatória da distribuição $\Gamma(\alpha, \beta)$. Encontre o estimador de α e β pelo método dos momentos.