

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

1 Método dos Momentos

2 Método da Máxima Verossimilhança

Introdução ao Método dos Momentos

O método dos momentos é uma técnica desenvolvida pelo matemático e estatístico Karl Pearson no final do século XIX. O método consiste em igualar os momentos amostrais de uma distribuição teórica com os momentos correspondentes da distribuição amostral, e estimar os parâmetros da distribuição teórica a partir desses momentos.

Introdução ao Método dos Momentos

O método dos momentos é uma técnica desenvolvida pelo matemático e estatístico Karl Pearson no final do século XIX. O método consiste em igualar os momentos amostrais de uma distribuição teórica com os momentos correspondentes da distribuição amostral, e estimar os parâmetros da distribuição teórica a partir desses momentos.

É amplamente aplicado em áreas como Engenharia, Ciências Sociais e Biológicas, dentre outras. É um dos mais simples e intuitivos métodos de inferência estatística, e pode ser facilmente adaptado a diferentes distribuições de probabilidade, o que o torna uma ferramenta útil na análise de dados.

Momentos

Antes de defini-lo formalmente vamos recordar o que são Momentos de Variáveis Aleatórias.

Antes de defini-lo formalmente vamos recordar o que são Momentos de Variáveis Aleatórias.

Definição 1

Para $k = 1, 2, \dots$, o momento de ordem k da variável X é definido por $E(X^k)$, desde que essa quantidade exista. Se $E(X) = \mu < \infty$, definimos o momento central de ordem k por $E[(X - \mu)^k]$, sempre que essa quantidade exista. De modo similar, o momento absoluto de ordem k da variável aleatória X é definido por $E(|X|^k)$.

Exemplo 1

Considerando $X \sim \Gamma(\alpha, \beta)$, calcule seus momentos.

$$\begin{aligned} E(X^k) &= \int_0^{\infty} x^k \cdot \frac{\beta^{\alpha} x^{\alpha-1} \cdot e^{-\beta x}}{\Gamma(\alpha)} dx \\ &= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha+k-1} \cdot e^{-\beta x} dx \end{aligned}$$

Sabe-se que

$$\int_0^{\infty} x^{p-1} \cdot e^{-qx} dx = \Gamma(p) \cdot q^{-p}$$

Exemplo 1

Aplicando essa propriedade à integral, obtemos:

$$\int_0^{\infty} x^{\alpha+k-1} \cdot e^{-\beta x} dx = \Gamma(\alpha + k) \cdot (\beta)^{-(\alpha+k)}$$

Substituindo na expressão original:

$$E(X^k) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \cdot \Gamma(\alpha + k) \cdot (\beta)^{-(\alpha+k)}$$

$$E(X^k) = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)} \cdot \beta^{-k} = \frac{\alpha(\alpha + 1) \cdots (\alpha + k - 1)}{\beta^k}$$

Esse é o valor esperado $E(X^k)$ em termos das funções Gama e do parâmetro β .

Exemplo 1

Observe que se $\alpha = 1$, X tem distribuição exponencial de parâmetro $\beta > 0$ e, assim,

$$E(X^k) = \frac{k!}{\beta^k}$$

Se fizermos $k = 1$, obtemos a média desta variável que é $\frac{1}{\beta}$.

Exemplo 2

Sabendo que $X \stackrel{\text{iid}}{\sim} B(n, p)$. Encontre o momento central de ordem 1 e 2 desta variável.

$$\begin{aligned} E(X^2) &= \sum_{i=0}^n i^2 \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=0}^n [i(i-1) + i] \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=2}^n i(i-1) \binom{n}{i} p^i (1-p)^{n-i} + E(X) \\ &= n(n-1) \sum_{i=2}^n \binom{n-2}{i-2} p^i (1-p)^{n-i} + E(X) \end{aligned}$$

Exemplo 2

$$\begin{aligned} E(X^2) &= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{(n-2)-j} + E(X) \\ &= n(n-1)p^2 [p + (1-p)]^{(n-2)} + np \\ &= n^2 p^2 + np(1-p) \end{aligned}$$

Exemplo 2

$$\begin{aligned}E(X^2) &= n(n-1)p^2 \sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{(n-2)-j} + E(X) \\&= n(n-1)p^2 [p + (1-p)]^{(n-2)} + np \\&= n^2 p^2 + np(1-p)\end{aligned}$$

Assim, $\text{Var}(X) = E(X - E(X))^2 = E(X^2) - (E(X))^2 = np(1-p)$

Método dos Momentos

Como já falado, o método dos momentos consiste em igualar os momentos amostrais de uma distribuição teórica com os momentos correspondentes da distribuição amostral, e estimar os parâmetros da distribuição teórica a partir desses momentos.

Método dos Momentos

Como já falado, o método dos momentos consiste em igualar os momentos amostrais de uma distribuição teórica com os momentos correspondentes da distribuição amostral, e estimar os parâmetros da distribuição teórica a partir desses momentos.

Seja

$$m_r = \frac{1}{n} \sum_{i=1}^n X_i^r, \quad r \geq 1 \quad (1)$$

o r -ésimo momento amostral de uma amostra aleatória X_1, \dots, X_n .

Seja,

$$\mu_r = E(X^r), \quad r \geq 1 \quad (2)$$

o r -ésimo momento populacional.

O método dos momentos consiste na obtenção de estimadores para $\theta = (\theta_1, \dots, \theta_k)$, resolvendo as equações

$$m_r = \mu_r, \quad r = 1, \dots, k. \quad (3)$$

$$\frac{1}{n} \sum_{i=1}^n X_i^r = E(X^r) \quad (4)$$

onde X_1, X_2, \dots, X_n são as observações amostrais, μ_r é o r -ésimo momento teórico da distribuição, e n é o tamanho da amostra. A partir dessa equação, é possível obter estimativas dos parâmetros da distribuição teórica por meio da solução de um sistema de equações, em que as equações correspondem aos primeiros momentos até a ordem k da distribuição.

Exemplo 1

Suponha que X_1, X_2, \dots, X_n sejam variáveis aleatórias independentes e identicamente distribuídas (iid) com distribuição $N(\theta, \sigma^2)$. Temos

$m_1 = \bar{X}$, $m_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$, $\mu'_1 = \theta$, $\mu'_2 = \theta^2 + \sigma^2$, e, portanto, devemos resolver

$$\bar{X} = \theta \text{ e } \frac{1}{n} \sum_{i=1}^n X_i^2 = \theta^2 + \sigma^2.$$

Resolvendo para θ e σ^2 , obtemos os estimadores pelo método de momentos:

$$\theta = \bar{X} \text{ e } \frac{1}{n} \sum_{i=1}^n X_i^2 = \theta^2 + \sigma^2 = \bar{X}^2 + \sigma^2 \Rightarrow \sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Exemplo 2

Sejam X_1, \dots, X_n uma amostra aleatória da distribuição de X , com densidade gama de parâmetros α e β dada por

$$f(x) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, \quad x > 0, \alpha > 0, \beta > 0.$$

Sabendo que

$$E[X] = \frac{\alpha}{\beta} \quad \text{e} \quad \text{Var}[X] = \frac{\alpha}{\beta^2},$$

obtemos que os estimadores para α e β são obtidos como solução das equações

$$\frac{\hat{\alpha}}{\hat{\beta}} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{e} \quad \frac{\hat{\alpha}^2}{\hat{\beta}^2} + \frac{\hat{\alpha}}{\hat{\beta}^2} = \frac{1}{n} \sum_{i=1}^n X_i^2$$

Exemplo 2

Logo,

$$\frac{\hat{\alpha}}{\hat{\beta}} = \bar{X} \Rightarrow \hat{\alpha} = \hat{\beta} \bar{X} \Rightarrow \hat{\alpha}^2 = \hat{\beta}^2 \bar{X}^2$$

$$\Rightarrow \frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{\hat{\alpha}^2 + \hat{\alpha}}{\hat{\beta}^2} = \frac{\hat{\beta}^2 \bar{X}^2 + \hat{\beta} \bar{X}}{\hat{\beta}^2} = \frac{\hat{\beta} \bar{X}^2 + \bar{X}}{\hat{\beta}}$$

$$\Rightarrow \frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{\bar{X}}{\hat{\beta}} + \bar{X}^2 \Rightarrow \hat{\beta} = \frac{\bar{X}}{\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2} = \frac{\bar{X}}{\hat{\sigma}^2}$$

$$\Rightarrow \hat{\alpha} = \frac{\bar{X}^2}{\hat{\sigma}^2}, \text{ em que } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=0}^n (X_i - \bar{X})^2$$

Introdução ao Método da Máxima Verossimilhança

O método de estimação por Máxima Verossimilhança foi criado pelo estatístico britânico Ronald A. Fisher em 1912. Sua finalidade é estimar os valores desconhecidos dos parâmetros de um modelo estatístico com base em dados observados, maximizando a verossimilhança dos dados. Isso significa encontrar os valores dos parâmetros que tornam os dados observados mais prováveis de terem sido gerados pelo modelo proposto.

As vantagens do método incluem sua simplicidade e robustez, bem como sua ampla aplicabilidade em diversas áreas da ciência, desde a física e a biologia até a economia e as ciências sociais. Além disso, o método de máxima verossimilhança é frequentemente utilizado como base para métodos mais avançados de inferência estatística, como a análise de variância e a regressão linear.

O que é Verossimilhança dos dados?

A verossimilhança dos dados é uma medida de quão provável é um determinado conjunto de dados ter sido gerado por um modelo estatístico específico. Em outras palavras, a verossimilhança representa a probabilidade de se obter os dados observados, dada uma hipótese sobre os parâmetros desconhecidos do modelo. Quanto maior a verossimilhança, mais provável é que os dados tenham sido gerados pelo modelo em questão. A estimação dos parâmetros do modelo por máxima verossimilhança envolve a busca pelos valores dos parâmetros que maximizam a verossimilhança dos dados.

Definição 2

Dada uma amostra X_1, X_2, \dots, X_n iid de uma população com fdp ou fp $f(x, \theta)$, em que θ é um parâmetro desconhecido. A base para nossos procedimentos inferenciais será a função de verossimilhança definida por:

$$L(\theta) = \prod_{i=1}^n f(x_i, \theta), \quad (5)$$

Em que, x_i é o valor observado de X_i , $i = 1, \dots, n$.

O logaritmo da função (5) é mais tratável matematicamente. Trabalharemos então com ele, que é conhecido como log-verossimilhança.

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

Exemplo

Sejam X_1, X_2, \dots, X_n uma amostra aleatória de uma distribuição com função de probabilidade

$$P(x) = \theta^x (1 - \theta)^{1-x}, \quad x \in \{0, 1\} \text{ e } \theta \in [0, 1].$$

Exemplo

Sejam X_1, X_2, \dots, X_n uma amostra aleatória de uma distribuição com função de probabilidade

$$P(x) = \theta^x (1 - \theta)^{1-x}, \quad x \in \{0, 1\} \text{ e } \theta \in [0, 1].$$

A função de verossimilhança é dada por:

$$L(\theta) = \prod_{i=1}^n P(X_i = x_i) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}.$$

O valor de θ que maximiza a função anterior é um bom estimador para θ . Ele dá a maior probabilidade para a amostra observada. Neste caso a função log-verossimilhança é dada por:

$$\begin{aligned}\ell(\theta) &= \sum_{i=1}^n x_i \log \theta + (n - \sum_{i=1}^n x_i) \log (1 - \theta) \\ \Rightarrow \frac{\partial \ell(\theta)}{\partial \theta} &= \frac{\sum_{i=1}^n x_i}{\theta} - \frac{n - \sum_{i=1}^n x_i}{1 - \theta}\end{aligned}$$

Para achar o ponto crítico, igualamos a derivada a zero, ou seja:

$$\frac{\partial \ell(\theta)}{\partial \theta} = 0 \Rightarrow \hat{\theta} = \frac{\sum_{i=1}^n x_i}{n}$$

é o estimador de máxima verossimilhança de θ .

Condições de Regularidade

Seja θ_0 o valor verdadeiro de θ . O próximo teorema nos dá uma justificativa teórica para maximizar a função de verossimilhança. Para isso assumiremos algumas condições de regularidade.

- (R0) Se $\theta \neq \theta'$, então $f(x, \theta) \neq f(x, \theta')$.
- (R1) As densidades possuem o mesmo suporte para todo θ , ou seja, o suporte não pode depender do parâmetro.
- (R2) O ponto θ_0 é um ponto interior de Ω , em que Ω é o suporte da função densidade.

Teorema 1

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

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

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

Demonstração

$$\begin{aligned} L(\theta_0, X) > L(\theta, X) &\Leftrightarrow \prod_{i=1}^n f(x_i, \theta_0) > \prod_{i=1}^n f(x_i, \theta) \\ &\Leftrightarrow \sum_{i=1}^n \log f(x_i, \theta_0) > \sum_{i=1}^n \log f(x_i, \theta) \\ &\Leftrightarrow \sum_{i=1}^n \log \left(\frac{f(x_i, \theta)}{f(x_i, \theta_0)} \right) < 0 \\ &\Leftrightarrow \frac{1}{n} \sum_{i=1}^n \log \left(\frac{f(x_i, \theta)}{f(x_i, \theta_0)} \right) < 0 \end{aligned}$$

Seja $y_i = \log \left(\frac{f(x_i, \theta)}{f(x_i, \theta_0)} \right)$, $i = 1, \dots, n$.

Continuação da Demonstração

Y_1, \dots, Y_n é uma sequência de v.a. iid com média finita. Pela lei fraca dos grandes números

$$\frac{\sum_{i=1}^n y_i}{n} \xrightarrow{P} E_{\theta_0}(Y_i) = E \left[\log \left(\frac{f(x_i, \theta)}{f(x_i, \theta_0)} \right) \right]$$
$$< \log E_{\theta_0} \left(\frac{f(x_i, \theta)}{f(x_i, \theta_0)} \right)$$

Desigualdade de Jensen

Continuação da Demonstração

Note que,

$$E_{\theta_0} \left(\frac{f(x_i, \theta)}{f(x_i, \theta_0)} \right) = \int_{\mathbb{R}} \frac{f(x_i, \theta)}{f(x_i, \theta_0)} f(x, \theta_0) dx = 1$$

f de θ e f de θ_0 tem o mesmo suporte

$$\Rightarrow \frac{\sum_{i=1}^n y_i}{n} < 0 \text{ para } n \rightarrow +\infty$$

$$\Rightarrow \frac{1}{n} \sum_{i=1}^n \log \left(\frac{f(x_i, \theta)}{f(x_i, \theta_0)} \right) \text{ converge em probabilidade}$$

para uma constante negativa. Assim, $L(\theta_0, X) > L(\theta, X)$ para $\theta \neq \theta_0$.

Definição 3

Dizemos que $\hat{\theta} = \hat{\theta}(x)$ é um Estimador de Máxima Verossimilhança (EMV) de θ se

$$\hat{\theta} = \operatorname{argmax}_{\theta} L(\theta, \mathbf{X}), \hat{\theta} \in \Theta.$$

Exemplo

Sejam $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \exp(\theta)$,

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, x > 0, \theta > 0$$

Exemplo

Sejam $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \exp(\theta)$,

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, x > 0, \theta > 0$$

A log-verossimilhança fica dada por

$$\begin{aligned}\ell(\theta) &= -n \log \theta - \theta^{-1} \sum_{i=1}^n x_i \\ \Rightarrow \hat{\theta} &= \bar{X}\end{aligned}$$

Exemplo

Sejam $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} U(0, \theta]$,

$$f(x) = \frac{1}{\theta}, x \in (0, \theta], \theta > 0$$

Exemplo

Sejam $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} U(0, \theta]$,

$$f(x) = \frac{1}{\theta}, x \in (0, \theta], \theta > 0$$

Para 

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n f(x_i, \theta) = \prod_{i=1}^n \frac{1}{\theta} \delta(0, \theta], \quad \delta(0, \theta] = \begin{cases} 1, & \text{se } x_i \in (0, \theta] \\ 0, & \text{c.c} \end{cases} \\ &= \left(\frac{1}{\theta}\right)^n \delta\{\max(x_1, \dots, x_n) \leq \theta\} \end{aligned}$$

Mostre que $\hat{\theta} = \max(X_1, \dots, X_n)$

Em alguns exemplos simples, a solução da equação de verossimilhança pode ser obtida explicitamente. Em situações mais complicadas, a solução da equação (6) será em geral obtida por procedimentos numéricos. Para se concluir que a solução da equação (6) é um ponto de máximo, é necessário verificar se:

$$\ell''(\hat{\theta}; \mathbf{x}) = \frac{\partial^2 \ell(\theta; \mathbf{x})}{\partial^2 \theta} \Big|_{\theta=\hat{\theta}} < 0. \quad (6)$$

Encontrar o EMV pode ser um problema complexo em alguns casos, devido a várias dificuldades, incluindo:

- **Função de verossimilhança complexa:** Em muitos casos, a função de verossimilhança é complexa e não possui uma forma analítica simples. Isso pode tornar a maximização da função de verossimilhança um problema difícil e pode exigir o uso de técnicas computacionais sofisticadas.

Encontrar o EMV pode ser um problema complexo em alguns casos, devido a várias dificuldades, incluindo:

- **Função de verossimilhança complexa:** Em muitos casos, a função de verossimilhança é complexa e não possui uma forma analítica simples. Isso pode tornar a maximização da função de verossimilhança um problema difícil e pode exigir o uso de técnicas computacionais sofisticadas.
- **Convergência:** O EMV é encontrado pela maximização da função de verossimilhança. Em alguns casos, a função de verossimilhança pode ter várias máximas locais, o que pode dificultar a convergência do algoritmo de otimização para a solução global. Isso pode ser especialmente problemático se a função de verossimilhança for multimodal e as várias máximas locais estiverem próximas em termos de valor.

- **Dados insuficientes:** Em alguns casos, a quantidade de dados disponíveis pode não ser suficiente para permitir uma estimativa precisa dos parâmetros do modelo. Isso pode tornar a estimativa do EMV imprecisa e pode resultar em intervalos de confiança amplos ou em estimativas enviesadas.

- **Dados insuficientes:** Em alguns casos, a quantidade de dados disponíveis pode não ser suficiente para permitir uma estimativa precisa dos parâmetros do modelo. Isso pode tornar a estimativa do EMV imprecisa e pode resultar em intervalos de confiança amplos ou em estimativas enviesadas.
- **Modelos mal especificados:** O EMV é um método que depende da escolha do modelo estatístico correto. Se o modelo estiver mal especificado, o EMV pode levar a estimativas imprecisas ou enviesadas dos parâmetros do modelo.

- **Computacionalmente caro:** Em alguns casos, a computação do EMV pode ser computacionalmente cara, especialmente se a função de verossimilhança for complexa e exigir muitos cálculos. Isso pode tornar o EMV impraticável em alguns casos.

- **Computacionalmente caro:** Em alguns casos, a computação do EMV pode ser computacionalmente cara, especialmente se a função de verossimilhança for complexa e exigir muitos cálculos. Isso pode tornar o EMV impraticável em alguns casos.

Em resumo, o EMV pode ser difícil de encontrar em algumas situações, especialmente quando a função de verossimilhança é complexa, a convergência é difícil, os dados são insuficientes, o modelo está mal especificado ou a computação é cara. No entanto, em muitos casos, o EMV é um método poderoso e útil para estimar os parâmetros do modelo.

Exemplo 1

Verossimilhança da Normal

Sejam $X_1, \dots, X_n \sim N(\theta, 1)$ iid e $L(\theta|\mathbf{x})$ denota a função de verossimilhança. Então,

$$L(\theta|\mathbf{x}) = \prod_{i=1}^n \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-(1/2)(x_i - \theta)^2} = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^n (x_i - \theta)^2}.$$

A equação $\frac{d}{d\theta} L(\theta|\mathbf{x}) = 0$ se reduz para

$$\sum_{i=1}^n (x_i - \theta) = 0$$

que tem solução $\hat{\theta} = \bar{x}$

Exemplo 1

Verossimilhança da Normal

Assim, \bar{x} é um candidato para o EMV. Para verificar que \bar{x} é um máximo global, observamos que ele é a única solução para $\sum_{i=1}^n (x_i - \theta) = 0$. Além disso, é possível observar que

$$\frac{d^2}{d\theta^2} L(\theta|\mathbf{x})|_{\theta=\bar{x}} < 0$$

e que $\lim_{\theta \rightarrow \pm\infty} L(\theta|\mathbf{x}) = 0$. Logo, \bar{x} é um máximo global, isto é, \bar{x} é o EMV de θ .

Exemplo 2

EMV de Bernoulli

Sejam X_1, \dots, X_n Bernoulli(p) iid. Então, a função de verossimilhança é,

$$L(p|\mathbf{x}) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^y (1-p)^{n-y}, \text{ em que, } y = \sum_{i=1}^n x_i.$$

Considere o log de verossimilhança dado por,

$$\ell(p|\mathbf{x}) = \log L(p|\mathbf{x}) = y \log p + (n-y) \log (1-p).$$

Exemplo 2

EMV de Bernoulli

Sejam X_1, \dots, X_n Bernoulli(p) iid. Então, a função de verossimilhança é,

$$L(p|x) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^y (1-p)^{n-y}, \text{ em que, } y = \sum_{i=1}^n x_i.$$

Considere o log de verossimilhança dado por,

$$\ell(p|x) = \log L(p|x) = y \log p + (n-y) \log (1-p).$$

EMV de Bernoulli

Se $0 < y < n$, diferenciar $\log L(p|x)$ e igualar o resultado a zero fornece a solução $\hat{p} = \frac{y}{n}$.

Exemplo 2

EMV de Bernoulli

É fácil verificar que $\hat{p} = \frac{y}{n}$ é máximo global, pois se $y = 0$ ou $y = n$, temos

$$\log L(p|\mathbf{x}) = \begin{cases} n \log(1 - p), & \text{se } y = 0 \\ n \log p, & \text{se } y = n \end{cases}.$$

Em qualquer um dos casos temos $\log L(p|\mathbf{x})$ é uma função monótona de p e, portanto, $\hat{p} = \frac{y}{n}$ é ponto de máximo. Assim, $\hat{p} = \frac{y}{n}$ é o EMV de p .

Teorema 2

Seja X_1, X_2, \dots, X_n uma a.a. de uma distribuição com densidade $f(\mathbf{x}; \theta)$, $\theta \in \Theta$. Seja g uma função especificada e $\eta = g(\theta)$. Suponha $\hat{\theta}$ o EMV de θ , então $\hat{\eta} = g(\hat{\theta})$ é o EMV de η .

Princípio da Invariância dos EMV

Em outras palavras, se estamos interessados em estimar alguma quantidade de interesse (η) que pode ser expressa como uma função dos parâmetros (θ) da distribuição, podemos aplicar o EMV aos parâmetros e, em seguida, aplicar a função g aos resultados para obter o estimador do parâmetro η . Isso é válido sob certas condições, e é uma propriedade muito útil da Estimativa de Máxima Verossimilhança, pois nos permite estimar uma ampla variedade de quantidades de interesse de maneira eficaz.

Princípio da Invariância dos EMV

Em outras palavras, se estamos interessados em estimar alguma quantidade de interesse (η) que pode ser expressa como uma função dos parâmetros (θ) da distribuição, podemos aplicar o EMV aos parâmetros e, em seguida, aplicar a função g aos resultados para obter o estimador do parâmetro η . Isso é válido sob certas condições, e é uma propriedade muito útil da Estimativa de Máxima Verossimilhança, pois nos permite estimar uma ampla variedade de quantidades de interesse de maneira eficaz.

Em resumo, o Princípio da Invariância dos Estimadores de Verossimilhança estabelece que aplicar uma função a um estimador de EMV resulta no estimador de EMV para a mesma função aplicada aos parâmetros originais.

Demonstração

Suponha g invertível. A verossimilhança de interesse é $L(g^{-1}(\eta))$. Esta função é maximizada quando $g^{-1}(\eta) = \hat{\theta}$, isto é, tome $\hat{\eta} = g(\hat{\theta})$.

Demonstração

Suponha g não invertível. Para cada valor de η dentro do alcance da função g , definimos o conjunto (pré-imagem) $g^{-1}(\eta) = \{\theta : g(\theta) = \eta\}$. O valor máximo de L ocorre em $\hat{\theta}$, e o domínio de g é Ω , abrangendo $\hat{\theta}$. Portanto, $\hat{\theta}$ pertence a um destes conjuntos de pré-imagem, e na verdade, ele pode pertencer apenas a uma única pré-imagem. Para maximizar a função $L(\eta)$, devemos selecionar $\hat{\eta}$ de tal forma que $g^{-1}(\hat{\eta})$ corresponda a essa pré-imagem única contendo $\hat{\theta}$. Isso leva à equação $\hat{\eta} = g(\hat{\theta})$.

Teorema 3

Assuma X_1, \dots, X_n satisfazendo as condições de regularidade $R0, R1$ e $R2$ com θ_0 sendo o verdadeiro valor de θ . Além disso, assuma que $f(x, \theta)$ é diferenciável com relação a $\theta \in \Theta$. Então, a equação

$$\frac{\partial L(\theta)}{\partial \theta} = 0, \text{ ou, equivalentemente, } \frac{\partial \ell(\theta)}{\partial \theta} = 0,$$

tem uma solução $\hat{\theta}_n$ tal que $\hat{\theta}_n \xrightarrow{P} \theta_0$.

Demonstração

Como θ_0 é um ponto interior de Θ , existe $a > 0$ tal que $(\theta_0 - a, \theta_0 + a) \subset \Theta$. Defina,

$$S_n = \{L(\theta_0, \mathbf{X}) > L(\theta_0 + a, \mathbf{X})\} \cap \{L(\theta_0, \mathbf{X}) > L(\theta_0 - a, \mathbf{X})\}$$

Provamos (Ver Teorema (1)) que $\lim_{n \rightarrow \infty} P(S_n) = 1$.

Em S_n temos a existência de um máximo local, digamos $\hat{\theta}_n$, ou seja, $\theta_0 - a < \hat{\theta}_n < \theta_0 + a$ e $\frac{\partial \ell(\theta)}{\partial \theta} \big|_{\theta=\hat{\theta}_n} = 0$. Portanto,

$$\begin{aligned} S_n &\subset \{|\hat{\theta}_n - \theta_0| < a\} \cap \left\{ \frac{\partial \ell(\theta)}{\partial \theta} = 0 \right\} \\ \Rightarrow P(S_n) &\leq P\left(\{|\hat{\theta}_n - \theta_0| < a\} \cap \left\{ \frac{\partial \ell(\theta)}{\partial \theta} = 0 \right\}\right) \end{aligned}$$

Continuação da Demonstração

$$P(S_n) \leq P\left(\{|\hat{\theta}_n - \theta_0| < a\} \cap \left\{\frac{\partial \ell(\theta)}{\partial \theta} = 0\right\}\right)$$

$$\Rightarrow \liminf_{n \rightarrow \infty} P(S_n) \leq \liminf_{n \rightarrow \infty} P\left(\{|\hat{\theta}_n - \theta_0| < a\} \cap \left\{\frac{\partial \ell(\theta)}{\partial \theta} = 0\right\}\right)$$

$$\Rightarrow \lim_{n \rightarrow \infty} P\left(\{|\hat{\theta}_n - \theta_0| < a\} \cap \left\{\frac{\partial \ell(\theta)}{\partial \theta} = 0\right\}\right) = 1$$

pois $\liminf_{n \rightarrow \infty} P(S_n) = 1$. Note que $P(\cdot) \geq 1 \Rightarrow P(\cdot) = 1$.

Para 

Exercícios 6.1.1(letra a), 6.1.2, 6.1.4, 6.1.6, 6.1.9, 6.1.10 e 6.1.12

Referências I

BOLFARINE, Heleno; SANDOVAL, Mônica Carneiro. **Introdução à inferência estatística**. [S.l.]: SBM, 2001. v. 2.

CASELLA, George; BERGER, Roger L. **Statistical inference**. [S.l.]: Cengage Learning, 2021.

HOGG, RV; MCKEAN, J; CRAIG, AT. **Introduction to Mathematical Statistics**. Eighth Edition. [S.l.]: Pearson, 2019.