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

Introdução ao Método dos Momentos

O método dos momentos é uma técnica desenvolvida pelo matemático e estatístico Karl Pearson no final do século XIX.

Introdução ao Método dos Momentos

O método dos momentos é uma técnica desenvolvida pelo matemático e estatístico Karl Pearson no final do século XIX.

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.

Método dos Momentos

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

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

$$m_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r, \; r \geq 1 \tag{1}$$

o $r-$ésimo momento amostral de uma amostra aleatória $X_1, \ldots, X_n$. Seja,

$$\mu_r = E(X^r), \; r \geq 1 \tag{2}$$

o $r-$ésimo momento populacional.

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

$$m_r = \mu_r, \; r = 1, \ldots, k. \tag{3}$$

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

onde $X_1, X_2, ..., X_n$ 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 $X_1, X_2, \ldots, X_n$ sejam variáveis aleatórias independentes e identicamente distribuídas (iid) com distribuição $N(\theta, \sigma^2)$. Temos $m_1 = \bar{X}$, $m_2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2$, $\mu_1' = \theta$, $\mu_2' = \theta^2 + \sigma^2$, 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 $X_1, \ldots, X_n$ 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

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

Exemplo 2

Logo,

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

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.

Definições Importantes

Definição 1

Dada uma amostra $X_1, X_2, ..., X_n$ *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:*

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

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

Log-Verossimilhança

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 $X_1, X_2, ..., X_n$ uma amostra aleatória de uma distribuição com função de probabilidade

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

Exemplo

Sejam $X_1, X_2, ..., X_n$ uma amostra aleatória de uma distribuição com função de probabilidade

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

A função de verossimilhança é dada por:

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

Exemplo

Sejam $X_1, X_2, ..., X_n$ uma amostra aleatória de uma distribuição com função de probabilidade

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

A função de verossimilhança é dada por:

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

Veja o gráfico de verossimilhança, cliquem aqui!

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:

$$\ell(\theta) = \sum_{i=1}^{n} x_i \log\theta + (n - \sum_{i=1}^{n} x_i)\log(1-\theta)$$

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

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 maximar 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

Teorema 1

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

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

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

Continuação da Demonstração

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

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

Desigualdade de Jensen

Continuação da Demonstração

Note que,

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

f de θ e f de θ_0 tem o mesmo suporte

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

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

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

Estimador de Máxima Verossimilhança

Definição 2

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

$$\hat{\theta} = argmax_{\theta} L(\theta, \boldsymbol{X}), \hat{\theta} \in \Theta.$$

Exemplo

Sejam $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \exp(\theta)$,

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

Exemplo

Sejam $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \exp(\theta)$,

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

A log-verossimilhança fica dada por

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

$$\Rightarrow \hat{\theta} = \bar{X}$$

Exemplo

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

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

Exemplo

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

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

Para ⌂

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

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

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:

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

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

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

A equação $\frac{d}{d\theta} L(\theta|\boldsymbol{x}) = 0$ 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

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

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

Princípio da Invariância dos EMV

Teorema 2

Seja $X_1, X_2, ..., X_n$ *uma a.a. de uma distribuição com densidade* $f(\boldsymbol{x};\theta)$, $\theta \in \Theta$ *e* $g : \Theta \rightarrow \tau$ *uma função sobrejetora, denote* $\eta = g(\theta)$. *Suponha* $\hat{\theta}$ *o EMV de* θ, *então* $\hat{\eta} = g(\hat{\theta})$ *é o EMV de* η *da mesma amostra considerada.*

Princípio da Invariância dos EMV

Teorema 2

Seja $X_1, X_2, ..., X_n$ *uma a.a. de uma distribuição com densidade* $f(\boldsymbol{x};\theta)$, $\theta \in \Theta$ *e* $g : \Theta \to \tau$ *uma função sobrejetora, denote* $\eta = g(\theta)$. *Suponha* $\hat{\theta}$ *o EMV de* θ, *então* $\hat{\eta} = g(\hat{\theta})$ *é o EMV de* η *da mesma amostra considerada.*

Interpretação:

Mesmo estando η e θ em espaços paramétricos, não necessariamente, iguais, a estimativa de $\eta = g(\theta)$ que torna a amostra observada mais provável é o valor obtido após a aplicação da função g no valor de $\hat{\theta}$ (EMV de θ) que torna a amostra observada mais provável de ser observada!

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, pode ser, por exemplo, uma transformação do parâmetro θ, como o quadrado, o logaritmo, ou qualquer outra função. Basta, simplesmente, aplicar essa função ao EMV do parâmetro. 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.

Observação

Considerando $g : \Theta \to \tau$ uma função sobrejetora que mapeia $\theta \in \Theta$ em $\eta = g(\theta) \in \tau$. A função de verossimilhança de η é a mesma que a função de verossimilhança de θ, uma vez que η é uma função de θ, isto é:

$$L(\eta \mid \boldsymbol{X}_n) = L(\theta \mid \boldsymbol{X}_n).$$

Exemplo: Invariância do EMV em Bernoulli

Seja $X_1, X_2, \ldots, X_n \sim$ Bernoulli(p), onde $p \in (0,1)$.
A função de verossimilhança em termos de p é:

$$L(p \mid \mathbf{X}_n) = p^{\sum_{i=1}^{n} X_i}(1-p)^{n-\sum_{i=1}^{n} X_i}.$$

Mudança de Parâmetro

Defina

$$\theta = p \quad \text{e} \quad \eta = g(\theta) = 1 - \theta.$$

Ou seja, η é a probabilidade de sair coroa.
Como $\eta = 1 - p$, temos $p = 1 - \eta$. Logo,

$$L(\eta \mid \mathbf{X}_n) = (1-\eta)^{\sum X_i} \eta^{n - \sum X_i}.$$

Comparação

Note que:

$$L(p \mid \mathbf{X}_n) = p^{\sum X_i}(1-p)^{n-\sum X_i},$$

$$L(\eta \mid \mathbf{X}_n) = (1-\eta)^{\sum X_i}\,\eta^{n-\sum X_i}.$$

É a mesma função, apenas reparametrizada.

Conclusão

A verossimilhança depende dos dados, mas não do rótulo do parâmetro.

$$L(\eta \mid \mathbf{X}_n) = L(\theta \mid \mathbf{X}_n).$$

O princípio da invariância do EMV garante que $\hat{\eta} = g(\hat{\theta})$. Vamos provar!

Demonstração

Considere a função de verossimilhança de θ para a amostra dada,

$$L(\theta, X_1, \cdots, X_n) = L(\theta, \boldsymbol{X}_n) = \prod_{i=1}^{n} f(\theta, \boldsymbol{X}_n).$$

Queremos maximizar $L(\eta, \boldsymbol{X}_n) = L(g(\theta), \boldsymbol{X}_n) = L^g_{\boldsymbol{X}_n}(\eta)$, função de verossimilhança induzida pela função g dada a amostra $X_1, \cdots, X_n$ sob o conjunto τ. Na verdade, queremos mostrar que $L^g_{\boldsymbol{X}_n}(\hat{\eta}) = L_{\boldsymbol{X}_n}(\hat{\theta})$.

Demonstração

Considere a função de verossimilhança de θ para a amostra dada,

$$L(\theta, X_1, \cdots, X_n) = L(\theta, \boldsymbol{X}_n) = \prod_{i=1}^{n} f(\theta, \boldsymbol{X}_n).$$

Queremos maximizar $L(\eta, \boldsymbol{X}_n) = L(g(\theta), \boldsymbol{X}_n) = L^g_{\boldsymbol{X}_n}(\eta)$, função de verossimilhança induzida pela função g dada a amostra $X_1, \cdots, X_n$ sob o conjunto τ. Na verdade, queremos mostrar que $L^g_{\boldsymbol{X}_n}(\hat{\eta}) = L_{\boldsymbol{X}_n}(\hat{\theta})$.

Para isso, é necessário entender que o EMV para $\eta = g(\theta)$ é

$$L^g_{\boldsymbol{X}_n}(\eta) = \max_{\theta \in A(\eta)} L_{\boldsymbol{X}_n}(\theta), \eta \in \tau,$$

em que $A(\eta) = \{\theta \in \Theta; g(\theta) = \eta\} = g^{-1}(\{\eta\})$.

Questão Central?

Por que maximizar $L(\theta)$ em $\theta \in \Theta$ nos dá o EMV de $\eta \in \tau$, dado que η e θ pertencem a espaços diferentes?

Explicação simplificada:

Ligação entre η e θ

- O parâmetro η é uma função de θ, ou seja, todo valor de η corresponde a algum valor de θ.
- A relação entre η e θ é determinística, dada pela função g. Isso significa que, se você souber θ, você automaticamente sabe o valor de η, pois $\eta = g(\theta)$.

Explicação simplificada:

Ligação entre η e θ

- O parâmetro η é uma função de θ, ou seja, todo valor de η corresponde a algum valor de θ.
- A relação entre η e θ é determinística, dada pela função g. Isso significa que, se você souber θ, você automaticamente sabe o valor de η, pois $\eta = g(\theta)$.

Maximização em η

- Para encontrar o EMV de η, você precisa maximizar a função de verossimilhança em termos de η.
- Como $\eta = g(\theta)$, maximizar a verossimilhança em η é equivalente a maximizar em θ, pois toda maximização sobre η pode ser traduzida em uma maximização sobre os θ's que produzem esse η.

Explicação simplificada:

Conjunto $A(\eta)$

- Para cada η, o conjunto $A(\eta) = \{\theta \in \Theta; g(\theta) = \eta\}$ é o conjunto de todos os valores de θ que produzem esse valor de η.
- Maximizar $L(\theta)$ sobre $\theta \in A(\eta)$ encontra o valor de θ dentro do conjunto $A(\eta)$ que maximiza a verossimilhança.
- Como $A(\eta)$ contém todos os θ's que produzem um dado η, a maximização sobre $A(\eta)$ nos dá o maior valor possível de verossimilhança para aquele η.
- Isso significa que a função de verossimilhança para η, $L^g_{\boldsymbol{X}_n}(\eta)$, é definida como o valor máximo de $L(\theta)$ sobre $A(\eta)$.

Explicação simplificada:

Conclusão

Embora $\eta \in \tau$ e $\theta \in \Theta$ sejam de espaços diferentes, a maximização da verossimilhança sobre η pode ser reduzida a uma maximização sobre θ, pois η é uma função de θ. Maximizar $L(\theta)$ sobre o conjunto de θ's que correspondem a um dado η nos dá a verossimilhança para η. Isso justifica a expressão:

$$L^g_{\boldsymbol{X}_n}(\eta) = \max_{\theta \in A(\eta)} L_{\boldsymbol{X}_n}(\theta),$$

onde $A(\eta)$ é o conjunto de pré-imagens de η sob g.

Continuando a Demonstração. . .

É claro que $L(\hat{\theta}, \boldsymbol{X}_n) \geq L(\theta, \boldsymbol{X}_n), \forall\theta \in \Theta$ e que $U_{\eta\in\tau} A(\eta) = \Theta$. Também é fato, já que g é sobrejetora, que existe pelo menos um $\tilde{\eta}$ tal que $\tilde{\eta} = g(\hat{\theta})$, assim, $\hat{\theta} \in A(\tilde{\eta})$. Vamos denotar $\tilde{\eta}$ por $\hat{\eta}$. Logo,

$$L^g_{\boldsymbol{X}_n}(\hat{\eta}) = \max_{\theta\in A(\hat{\eta})} L(\theta) \geq L(\hat{\theta}, \boldsymbol{X}_n) = L_{\boldsymbol{X}_n}(\hat{\theta}),$$

pois $\hat{\theta} \in A(\hat{\eta})$. Aqui, estou restrito a $A(\hat{\eta})$.

Por que não podemos assumir diretamente que $L^g_{\boldsymbol{X}_n}(\hat{\eta}) = L_{\boldsymbol{X}_n}(\hat{\theta})$?

Aqui está o ponto crucial: embora $\hat{\theta} \in A(\hat{\eta})$, pode haver outros $\theta' \in A(\hat{\eta})$ que produzem o mesmo valor de $\hat{\eta}$, e esses θ' podem ter um valor de verossimilhança $L(\theta')$ maior ou igual a $L(\hat{\theta})$. Isso ocorre porque a função $g(\theta)$ **pode ser não injetora**, ou seja, pode haver **vários** θ's **diferentes** que correspondem ao mesmo valor de η. Se houver algum $\theta' \in A(\hat{\eta})$ com $L(\theta') > L(\hat{\theta})$, então:

$$L^g_{\boldsymbol{X}_n}(\hat{\eta}) = \max_{\theta \in A(\hat{\eta})} L_{\boldsymbol{X}_n}(\theta) \geq L_{\boldsymbol{X}_n}(\hat{\theta}),$$

porque a maximização sobre o conjunto $A(\hat{\eta})$ pode resultar em um valor de verossimilhança maior do que $L(\hat{\theta})$, que foi obtido ao maximizar sobre todo o espaço Θ, e não apenas sobre $A(\hat{\eta})$. Mas, lembre-se que $\theta' \in A(\hat{\eta})$.

Continuando a Demonstração. . .

Por outro lado, para $\eta \in g(\Theta) = \tau$,

$$L^g_{\boldsymbol{X}_n}(\eta) \leq \max_{\eta \in \tau} L^g_{\boldsymbol{X}_n}(\eta) \leq \max_{\eta \in \tau} \max_{\theta \in A(\eta)} L_{\boldsymbol{X}_n}(\theta)$$

$$\leq \max_{\theta \in \Theta} L_{\boldsymbol{X}_n}(\theta) = L_{\boldsymbol{X}_n}(\hat{\theta})$$

$$\Rightarrow L_{\boldsymbol{X}_n}(\eta) \leq L_{\boldsymbol{X}_n}(\hat{\theta}), \forall \eta \in \tau.$$

Demonstração

Logo,

$$L^g_{\boldsymbol{X}_n}(\hat{\eta}) = L_{\boldsymbol{X}_n}(\hat{\theta}) \Rightarrow \hat{\eta} = g(\hat{\theta}).$$

Exemplo

Considere $X_1, \ldots, X_n$ amostra aleatória de X ~Poisson(μ), $\mu \geq 0$. Encontre o EMV para $\sqrt{\mu}$.

Exemplo

Considere $X_1, \ldots, X_n$ amostra aleatória de X ~Poisson(μ), $\mu \geq 0$. Encontre o EMV para $\sqrt{\mu}$.

Sabemos que $L(\mu, \boldsymbol{X}_n) = \dfrac{\mu^{\sum\limits_{i=1}^{n} X_i} e^{-n\mu}}{\prod\limits_{i=1}^{n} X_i!}$ e que $\hat{\mu} = \bar{X}$. Pelo principio da invariância se $\eta = \sqrt{\mu} \Rightarrow \hat{\eta} = \sqrt{\hat{\mu}} = \sqrt{\bar{X}}$

Exemplo

Considere $X_1, \ldots, X_n$ amostra aleatória de X ~Poisson(θ), $\theta \geq 0$. Encontre o EMV para $P_\theta(X = k), k \in \mathbb{N}$.

Exemplo

Considere $X_1, \dots, X_n$ amostra aleatória de $X \sim$Poisson(θ), $\theta \geq 0$. Encontre o EMV para $P_\theta(X = k), k \in \mathbb{N}$.

Sabemos que $\hat{\mu} = \bar{X}$. Pelo principio da invariância se $\eta = P_\theta(X = k) = \dfrac{e^{-\theta}\theta^k}{k!}$, então

$$\hat{\eta} = g(\bar{X}) = \frac{e^{-\bar{X}}\bar{X}^k}{k!}$$

Exemplo

Considere $X_1, \ldots, X_n$ amostra aleatória de $X \sim \text{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$. Encontre o EMV para $CV_\theta(X) = \dfrac{\sqrt{\sigma^2}}{\mu}, \mu \neq 0.$

Exemplo

Considere $X_1, \ldots, X_n$ amostra aleatória de $X \sim \mathrm{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$. Encontre o EMV para $CV_\theta(X) = \dfrac{\sqrt{\sigma^2}}{\mu}, \mu \neq 0$.

Sabemos que $\hat{\theta} = \left(\bar{X}, \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)$. Pela invariância do EMV, temos que:

$$g(\hat{\theta}) = CV_{\hat{\theta}}(X) = \frac{\sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}}{\bar{X}}$$

é o EMV de $CV_\theta(X) = \dfrac{\sqrt{\sigma^2}}{\mu}, \mu \neq 0$.

Teorema

Teorema 3

Assuma $X_1, \ldots, X_n$ *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,

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

Provamos (Ver Teorema (1)) que $\lim\limits_{n \to \infty} P(S_n) = 1$.

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 $\dfrac{\partial \ell(\theta)}{\partial \theta}\Big|_{\theta = \hat{\theta}_n} = 0$. Portanto,

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

Continuação da Demonstração

$$
\begin{aligned}
P(S_n) &\leq P\left(\{|\hat{\theta}_n - \theta_0| < a\} \cap \left\{\frac{\partial \ell(\theta)}{\partial \theta} = 0\right\}\right) \\
\Rightarrow \liminf_{n\to\infty} P(S_n) &\leq \liminf_{n\to\infty} P\left(\{|\hat{\theta}_n - \theta_0| < a\} \cap \left\{\frac{\partial \ell(\theta)}{\partial \theta} = 0\right\}\right) \\
&\Rightarrow \lim_{n\to\infty} P\left(\{|\hat{\theta}_n - \theta_0| < a\} \cap \left\{\frac{\partial \ell(\theta)}{\partial \theta} = 0\right\}\right) = 1
\end{aligned}
$$

pois $\liminf_{n\to\infty} P(S_n) = 1$. 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 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*.