



# IND 1115

## Inferência Estatística

### Aula 10

Outubro 2004

Mônica Barros

monica@mbarros.com

1



## Conteúdo

- Método de Máxima Verossimilhança
- Invariância do MLE
- Estimador não tendencioso
- Estimador Consistente
- Erro Quadrático Médio (MSE = Mean Squared Error)”

monica@mbarros.com

2

## Método da Máx. Verossimilhança



- Exemplo 1 (Bernoulli)
- Sejam  $X_1, X_2, \dots, X_n$  iid Bernoulli( $\theta$ ).
- A função de probabilidade de cada  $X_i$  é:

$$f(x_i, \theta) = \theta^{x_i} (1 - \theta)^{1-x_i} \quad \theta \in (0,1), x_i = 0,1$$

- A função de verossimilhança o produto das funções de probabilidade individuais, isto é:

$$L(\theta) = f(x_1, x_2, \dots, x_n, \theta) = \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1-x_i} = \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i} =$$

$$= \theta^{n\bar{X}} (1 - \theta)^{n - n\bar{X}} = \exp\{n\bar{X} \cdot \log \theta + (n - n\bar{X}) \cdot \log(1 - \theta)\}$$

monica@mbarros.com

3

## Método da Máx. Verossimilhança



- A log verossimilhança é:

$$l(\theta) = \log(L(\theta)) = (\sum x_i) \log \theta + (n - \sum x_i) \log(1 - \theta) =$$

$$= n\bar{X} \cdot \log \theta + (n - n\bar{X}) \log(1 - \theta)$$

- Resolvendo a equação de verossimilhança leva a:

$$\frac{dl}{d\theta} = 0 \Rightarrow \frac{n\bar{X}}{\theta} - \frac{(n - n\bar{X})}{1 - \theta} = 0$$

$$\Leftrightarrow (1 - \theta)n\bar{X} = n\theta - n\bar{X}\theta \Leftrightarrow n\theta = n\bar{X}$$

- E então o MLE para  $\theta$  é a média amostral.
- Verifique que  $\left. \frac{d^2 l}{d\theta^2} \right|_{\theta = \bar{X}} < 0$
- de tal forma que a média amostral realmente **MAXIMIZA** a verossimilhança.

monica@mbarros.com

4

## Método da Máx. Verossimilhança



- **Exemplo 2 (Normal)**
- Sejam  $X_1, X_2, \dots, X_n$  iid Normal( $\mu, 1$ ), ou seja, uma Normal com média desconhecida e variância conhecida (e suposta igual a um sem perda de generalidade).
- Mostre que o MLE de  $\mu$  é a média amostral.

## Método da Máx. Verossimilhança



- A verossimilhança é:

$$L(\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(X_i - \mu)^2\right] = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2} \sum_{i=1}^n (X_i - \mu)^2\right\}$$

- Note que a verossimilhança é máxima quando  $Q(\mu) = \sum (X_i - \mu)^2$  é mínimo.
- Então é equivalente maximizar  $L(\mu)$  ou minimizar  $Q(\mu)$ .

## Método da Máx. Verossimilhança



$$\begin{aligned} Q(\mu) &= \sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i^2 - 2\mu X_i + \mu^2) = \\ &= \sum_{i=1}^n X_i^2 - 2\mu n\bar{X} + n\mu^2 \end{aligned}$$

- Derivando  $Q(\mu)$  em relação a  $\mu$  e igualando a zero nos leva a um ponto crítico:

$$\frac{dQ(\mu)}{d\mu} = 0 \rightarrow -2 \sum_{i=1}^n X_i + 2n\mu = 0 \Leftrightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

- Logo, o MLE de  $\mu$  é a média amostral.

## Método da Máx. Verossimilhança



- **Exemplo 3 (Normal com média e variância desconhecidas)**
- Sejam  $X_1, X_2, \dots, X_n$  iid Normal( $\mu, \sigma^2$ ), onde ambas a média e a variância são desconhecidas.
- Encontre os MLEs de  $\mu$  e  $\sigma^2$ .

### Exemplo 3



□ A verossimilhança é:

$$L(\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(X_i - \mu)^2\right] = \\ = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2\right\}$$

□ A log-verossimilhança é:

$$l(\mu) = \log L(\mu) = -\frac{n}{2} \log(2\pi\sigma^2) + \left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2\right\} = \\ = -\frac{n}{2} \log(2\pi\sigma^2) + \left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i^2 - 2\mu X_i + \mu^2)\right\}$$

### Exemplo 3



□ Derivando a log-verossimilhança com relação a  $\mu$  e igualando a zero temos:

$$\frac{dl(\mu)}{d\mu} = \frac{d}{d\mu} \left\{ -\frac{n}{2} \log(2\pi\sigma^2) + \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i^2 - 2\mu X_i + \mu^2) \right\} \right\} = \\ = 0 - \frac{1}{2\sigma^2} \frac{d}{d\mu} \left\{ \sum_{i=1}^n (X_i^2 - 2\mu X_i + \mu^2) \right\} = -\frac{1}{2\sigma^2} \left\{ -2 \sum_{i=1}^n X_i + 2n\mu \right\} \\ \Rightarrow \frac{dl(\mu)}{d\mu} = 0 \Leftrightarrow -2 \sum_{i=1}^n X_i + 2n\mu = 0 \Leftrightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

### Exemplo 3



□ Para encontrar o MLE de  $\sigma^2$  precisamos derivar a log-verossimilhança com relação a  $\sigma^2$  e igualar o resultado a zero. Note que é mais conveniente derivar em relação a  $\sigma^2$  (ao invés de  $\sigma$ ), o que evita sucessivas aplicações da regra da cadeia.

### Exemplo 3



$$\frac{dl(\sigma^2)}{d\sigma^2} = \frac{d}{d\sigma^2} \left\{ -\frac{n}{2} \log(2\pi\sigma^2) + -\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i^2 - 2\mu X_i + \mu^2) \right\} = \\ = -\frac{n}{2} \left( \frac{2\pi}{2\pi\sigma^2} \right) - \left\{ \frac{-1}{\sigma^4} \right\} \frac{\sum_{i=1}^n (X_i^2 - 2\mu X_i + \mu^2)}{2} = \\ = -\frac{n}{2} \left( \frac{1}{\sigma^2} \right) + \left\{ \frac{1}{2\sigma^4} \right\} \sum_{i=1}^n (X_i^2 - 2\mu X_i + \mu^2) \\ \Rightarrow \frac{dl(\sigma^2)}{d\sigma^2} = 0 \Leftrightarrow \frac{n}{2} \left( \frac{1}{\sigma^2} \right) = \left\{ \frac{1}{2\sigma^4} \right\} \sum_{i=1}^n (X_i^2 - 2\mu X_i + \mu^2) \\ \Leftrightarrow \frac{n}{\sigma^2} = \frac{\sum_{i=1}^n (X_i^2 - 2\mu X_i + \mu^2)}{\sigma^4} \Leftrightarrow \sigma^2 = \frac{\sum_{i=1}^n (X_i^2 - 2\mu X_i + \mu^2)}{n} = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$$

### Exemplo 3



- Mas, da última expressão, percebemos que o MLE de  $\sigma^2$  depende de  $\mu$ , que também é desconhecido. O que fazer? SUBSTITUIR  $\mu$  por seu MLE na equação, o que leva a:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \hat{\mu})^2}{n} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

- Logo, o MLE de  $\sigma^2$  é a **variância amostral**, mas com **denominador  $n$**  ao invés de  $(n-1)$ . Chamaremos este estimador de  $S^{*2}$  para diferenciá-lo da variância amostral definida anteriormente (com denominador  $n-1$ )

### Método da Máx. Verossimilhança



- Exemplo 4 (Uniforme)
- A densidade Uniforme é potencialmente problemática por que o **intervalo de valores possíveis da variável aleatória pode depender dos mesmos parâmetros desconhecidos** que se pretende estimar.
- Na distribuição Uniforme o(s) MLE(s) dos parâmetros desconhecidos **NÃO SÃO** obtidos pela derivada da log-verossimilhança? Por que?

### Exemplo 4



- Esta dificuldade surge porque os parâmetros da Uniforme são os extremos do intervalo de definição, e precisamos garantir que todas as observações da amostra estão dentro deste intervalo (no qual um ou os dois limites são desconhecidos), pois do contrário a verossimilhança será nula.

- Considere uma amostra aleatória da densidade  $\text{Unif}(\theta, 2)$ . A densidade de cada  $X_i$  é:

$$f(x_i, \theta) = \frac{1}{2-\theta} \quad \text{se } \theta \leq x_i \leq 2 \quad \text{e}$$

$$f(x_i, \theta) = 0 \quad \text{se } x_i \text{ está fora deste intervalo}$$

### Exemplo 4



- A verossimilhança é:

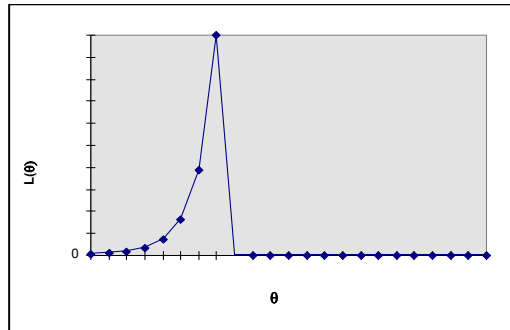
$$L(\theta) = \begin{cases} \frac{1}{(2-\theta)^n} & \text{se } \theta \leq x_1, x_2, \dots, x_n \leq 2 \\ 0 & \text{do contrário} \end{cases}$$

- Para que a verossimilhança seja não nula,  $\theta$  deve ser menor que todas as observações, e, em particular, menor que o mínimo da amostra, que é a primeira estatística de ordem,  $X_{(1)}$ .

## Exemplo 4



- O gráfico da verossimilhança é:



- Pelo gráfico, é óbvio que o valor de  $\theta$  que maximiza a verossimilhança é  $\hat{\theta} = X_{(1)}$ .

monica@mbarros.com

17

## Para Casa (Distribuição Uniforme)



- Sejam  $X_1, X_2, \dots, X_n$  iid  $\text{Unif}(0, \theta)$  onde  $\theta$  é desconhecido. Mostre que o MLE de  $\theta$  é  $X_{(n)} = \max(X_1, X_2, \dots, X_n)$ .
- Alguma modificação será necessária se a amostra for  $\text{Unif}(a, \theta)$  onde  $a$  é uma constante conhecida?

monica@mbarros.com

18

## Exemplo 5 (simulação da Uniforme)



- Gerei uma amostra aleatória de 50 valores de uma densidade Uniforme. Encontrei os seguintes valores:

|        |        |
|--------|--------|
| média  | 0.9136 |
| mínimo | 0.0290 |
| máximo | 1.9825 |

- O que você pode deduzir a partir destas informações? O intervalo de definição da densidade,  $(a,b)$ , é tal que  $a \leq 0.0290$  e  $b \geq 1.9825$ .

monica@mbarros.com

19

## Exemplo 5 (simulação da Uniforme)



- Agora eu gero uma amostra da mesma distribuição, mas com tamanho 10000 e encontro:

|        |        |
|--------|--------|
| média  | 1.0061 |
| mínimo | 0.0002 |
| máximo | 1.9996 |

- Note que o mínimo e o máximo estão se aproximando de 0 e 2 respectivamente, e estes foram os valores dos extremos da densidade Uniforme para a qual gerei as duas amostras!
- E se eu gerasse 1 milhão de observações, o que você acha que iria acontecer?

monica@mbarros.com

20

## Invariância do MLE



- ❑ Em muitos casos não estamos interessados em **estimar**  $\theta$ , e sim uma certa **função injetora**  $g(\cdot)$  deste parâmetro.
- ❑ O próximo teorema nos diz que o estimador de máxima verossimilhança para  $g(\theta)$  é apenas  $g(\cdot)$  aplicado ao estimador de máxima verossimilhança de  $\theta$ .

## Invariância do MLE



- ❑ Teorema (Princípio da invariância de um MLE)
- ❑ Considere uma amostra aleatória da densidade  $f(x, \theta)$  e seja  $L(\theta)$  a função de verossimilhança.
- ❑ Seja  $\hat{\theta}$  o estimador de máxima verossimilhança de  $\theta$  e suponha que queremos encontrar o MLE para  $g(\theta)$ , onde  $g(\cdot)$  é uma **função injetora** conhecida.
- ❑ Então o estimador de máxima verossimilhança para  $g(\theta)$  é  $g(\hat{\theta})$ , onde  $\hat{\theta}$  é o estimador de máxima verossimilhança para  $\theta$ .

## Invariância do MLE



- ❑ Demonstração
- ❑ Será omitida. *Na verdade este teorema é válido mesmo que  $g(\cdot)$  não seja injetora.*
- ❑ O princípio da invariância dos estimadores de máxima verossimilhança é muito importante, pois nos permite encontrar o MLE de qualquer função de  $\theta$  apenas através da aplicação desta função ao MLE de  $\theta$ .

## Exemplo 6 - MLE (Poisson)



- ❑ Considere a situação do exemplos da distribuição Poisson. O estimador de máxima verossimilhança de  $\theta$  é  $\bar{X} = 2$
- ❑ Qual o estimador de máxima verossimilhança de  $\theta^2, \sqrt{\theta}$  e  $\log \theta$  ?
- ❑ Qual o estimador de máxima verossimilhança de  $\Pr(X_1 \leq 1) = e^{-\theta} + \theta \cdot e^{-\theta} = e^{-\theta} \cdot (1 + \theta)$ ?

## Exemplo 6 - MLE (Poisson)



□ Pelo princípio da invariância dos estimadores de máxima verossimilhança, os MLEs são:

- o de  $\theta^2$  é 4,
- o de  $\sqrt{\theta}$  é 2,
- o de  $\log \theta$  é  $\log 2$  e o de  $\Pr(X_1 \leq 1)$  é:

$$e^{-\bar{X}}(1 + \bar{X}) = e^{-2}(1 + 2) = 0.406$$

## Estimador não tendencioso



□ A primeira característica desejável num estimador é a não tendenciosidade. Esta é uma característica desejável, mas **não é a única desejável** ou sequer a mais importante.

□ **Definição (Estimador não tendencioso)**

□ Seja  $\tilde{\theta} = T(X_1, X_2, \dots, X_n)$  um estimador para o parâmetro  $\theta$  de uma densidade  $f(x, \theta)$ .  $\tilde{\theta}$  é chamado de **não tendencioso** se  $E(\tilde{\theta}) = \theta$ , do contrário  $\tilde{\theta}$  é dito tendencioso.

## Exemplo 6 - MLE (Poisson)



□ Considere novamente os exemplos da distribuição de Poisson.

□ O estimador de máxima verossimilhança de  $\theta$  é  $\hat{\theta} = \bar{X}$

□ Mas, os  $X_i$  são iid Poisson( $\theta$ ), para  $i = 1, 2, \dots, n$ , e então  $E(X_i) = \theta$ . Mas,

$$\hat{\theta} = \bar{X} \text{ e } E(\hat{\theta}) = E(\bar{X}) = E(X_i) = \theta$$

e assim  $\hat{\theta} = \bar{X}$  é **não tendencioso** para  $\theta$ .

## Exemplo 7 - MLE (Normal)



□ Sejam  $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$ .

□ Sejam  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  e  $S^{*2} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$

□ Mas,

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

□ e então  $\bar{X}$  é um estimador **não tendencioso** para  $\mu$ .

□ Também,

$$E(S^{*2}) = \frac{(n-1)\sigma^2}{n} \neq \sigma^2$$

□ e assim  $S^{*2}$  é um estimador **tendencioso** para  $\sigma^2$ .

## Estimador Consistente



- ❑ Qualquer estatística usada para estimar  $\theta$  que converge estocasticamente para  $\theta$  é chamada um *estimador consistente* de  $\theta$ .
- ❑ Em palavras, à medida que o tamanho da amostra cresce, a probabilidade de um estimador consistente estar “longe” de  $\theta$  vai para zero (alternativamente, a probabilidade dele estar “perto” de  $\theta$  vai para 1).
- ❑ Consistência é uma propriedade desejável num estimador, e  muitas vezes o MLE é consistente.

## Estimador Consistente



- ❑ Provar que um estimador é consistente normalmente não é fácil, pois temos que estabelecer a convergência estocástica do estimador.
- ❑ Consistência é uma propriedade bastante forte, pois garante que um estimador  $\tilde{\theta}$  de um parâmetro  $\theta$  torna-se arbitrariamente próximo de  $\theta$  à medida que aumentamos o tamanho da amostra usado no cálculo de  $\tilde{\theta}$ .

## Estimador Consistente



- ❑ Por exemplo,  $\tilde{\theta} = \bar{X}$  é, pela lei fraca dos grandes números, um estimador consistente da média  $\mu$  da distribuição suposta para os dados, desde que a variância da distribuição seja finita.
- ❑ Assim, no exemplo da distribuição Poisson sabemos que o estimador de máxima verossimilhança de  $\theta$  tem 2 propriedades desejáveis: é não tendencioso e consistente.
- ❑ Na verdade, este resultado para a média amostral pode ser estendido.

## Estimador Consistente



- ❑ Considere uma amostra aleatória de uma distribuição qualquer com média  $\mu$  e variância finita. Então, pela lei fraca dos grandes números:

$$\bar{X} \xrightarrow{p} \mu$$

- ❑ Assim,  $\bar{X}$  é um *estimador consistente* de  $\mu$  para qualquer distribuição com média  $\mu$  e variância finita.

## Erro Quadrático Médio (MSE = Mean Squared Error)



□ Suponha que  $T = T(X_1, X_2, \dots, X_n)$  é um estimador de  $\tau(\theta)$  onde  $\tau(\theta)$  é uma função qualquer do parâmetro de interesse  $\theta$ .

□ O erro quadrático médio do estimador T é definido como:

$$MSE(T) = E[(T - \tau(\theta))^2]$$

□ Usando as definições anteriores, T é não tendencioso para  $\tau(\theta)$  se  $E(T) = \tau(\theta)$ . Também, T é um estimador consistente de  $\tau(\theta)$  se:

$$T \xrightarrow{p} \tau(\theta)$$

## Erro Quadrático Médio (MSE = Mean Squared Error)



□ Teorema

Se T é **não tendencioso** para  $\tau(\theta)$  então:

□ i)  $MSE(T) = VAR(T)$

□ Se T é **tendencioso** para  $\tau(\theta)$  então:

□ ii)  $MSE(T) = VAR(T) + (BIAS(T))^2$

onde  $BIAS(T) = E(T) - \tau(\theta)$  é chamado de "bias" ou tendência do estimador.

## Erro Quadrático Médio (MSE = Mean Squared Error)



□ Este resultado nos fornece uma maneira fácil de testar se um estimador é consistente.

□ Suponha que T é um estimador de  $\tau(\theta)$  baseado numa amostra de tamanho n.

□ Se **BIAS(T) e VAR(T) vão a zero** quando n tende a infinito, então T é **consistente**, pois isso implica no MSE(T) indo para zero quando n tende a infinito.

□ Se  $\lim BIAS(T) = 0$ , dizemos que T é assintoticamente não tendencioso.