Distribuições Amostrais e Estimação Pontual de Parâmetros - parte I

Marcos Oliveira Prates

2012/02

Objetivos

Ao final deste capítulo você deve ser capaz de:

- Entender estimação de parâmetros de uma distribuição de probabilidade ou de uma população.
- Explicar o papel da distribuição normal como distribuição amostral.
- Entender o Teorema Central do Limite e como usá-lo nesse caso.
- Entender o que são estimadores pontuais.

Introdução

- Inferência estatística:
  - tirar conclusões sobre a **população** a partir de uma **amostra**.
- Pode ser dividida em:
  - estimação de parâmetros;
  - testes de hipóteses.

Estimação de Parâmetros

- Um engenheiro analisa a resistência à tração de um componente de um automóvel.
- A resistência pode variar por motivos como:
  - diferença na matéria prima, variação no processo de fabricação e no processo de medida.
- O engenheiro precisa estimar a resistência dos componentes.
- Usa dados amostrais $\Rightarrow$ encontrar um valor razoável para a média verdadeira.
- Esse número é chamado **estimativa pontual**.

Testes de Hipótese

- Duas temperaturas, t_1 e t_2, podem ser usadas em um processo químico.
- O engenheiro suspeita que a temperatura t_1 resulta em rendimentos maiores.
- O teste estatístico resolve problemas desse tipo.
- A hipótese é de que:
  - o rendimento médio usando t_1 é maior que usando t_2.
- Não estamos interessados na estimação dos rendimentos.
- Queremos tirar conclusões acerca das **hipóteses** estabelecidas.

- Suponha que queremos estimar o parâmetro de uma população.
- Antes da coleta, os dados são variáveis aleatórias

$$X_1,\ X_2,\ \ldots,\ X_n\ .$$

- Qualquer função desses dados é também uma variável aleatória.
- Essa função é chamada de uma **estatística**.

Estatística

Qualquer função dos dados. Exemplo: média amostral.

- A estatística é uma variável aleatória $\Rightarrow$ possui distribuição de probabilidade.
- Chamamos a distribuição de uma estatística de **distribuição amostral.**
- A noção de distribuição amostral é muito importante na inferência estatística.

- Precisamos de um símbolo para denotar o parâmetro de interesse.
- Usaremos o símbolo θ para denotar o parâmetro de interesse.
- Exemplo: média de idade dos alunos da UFMG.
- Objetivo da estimação pontual:
  - selecionar um número (com base na amostra) que seja o valor mais plausível para θ.

- Seja X uma variável aleatória com função densidade $f(x)$.
- Essa função tem um parâmetro desconhecido θ.
- Seja
$$X_1,\ X_2,\ \ldots,\ X_n$$
uma amostra dessa variável.
- A estatística
$$\hat{\theta} = h(X_1,\ X_2,\ \ldots,\ X_n)$$
é um estimador pontual de θ.
- Depois de observada a amostra, $\hat{\theta}$ assume um valor numérico.
- Esse valor particular de $\hat{\theta}$ é chamado **estimativa**.

Estimador Pontual

- A estatística (**variável aleatória**) $\hat{\theta}$ é chamada **estimador pontual**.
- O valor observado de $\hat{\theta}$ é um **número** e é chamado de **estimativa pontual.**

Exemplo:

- Seja X uma variável aleatória normalmente distribuída.
- Suponha que a média μ de X não seja conhecida.
- A média amostral é um estimador de μ

$$\hat{\mu} = \bar{X} \, .$$

- Depois da amostra ter sido observada, o valor numérico de $\bar{x}$ é uma estimativa pontual de μ.
- Suponha que a amostra observada foi

$$x_1 = 25, \; x_2 = 30, \; x_3 = 29, \; x_4 = 31 \, .$$

- A estimativa de μ é

$$\bar{x} = \frac{25 + 30 + 29 + 31}{4} = 28,75 \, .$$

Algumas estimativas pontuais razoáveis:

- Para a média $\mu \Rightarrow \hat{\mu} = \bar{x}$.
- Para a variância $\sigma^2 \Rightarrow \hat{\sigma^2} = s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$.
- Para a proporção $p \Rightarrow$

$$\hat{p} = \frac{x}{n}$$

onde x é o número de ítens que pertencem à classe de interesse e n é o tamanho da amostra.

- Temos várias possibilidades de escolha de estimador de um parâmetro.
- Exemplo: a média da população pode ser estimada por:
  - média, mediana, ponto médio.
- Precisamos de critérios para decidir qual estimador é melhor.
- Isso vai depender das propriedades estatísticas do estimador.

Distribuições Amostrais e Teorema do Limite Central

- Inferência estatística:
  - tomar decisões sobre a população a partir da amostral.
- Podemos estar interessados no volume médio de enchimento de uma lata de refrigerante.
- O volume médio requerido é 300 mililitros.
- Um engenheiro retira uma amostra de 25 latas.
- O volume médio observado é $\bar{x} = 298$.
- O engenheiro decide que a média da população é 300 mililitros.
- Mesmo que a média da população seja 300 mililitros é muito provável que apareça uma amostra com média 298.

Ligação entre os modelos de probabilidade e os dados:

- O valor numérico dos dados é o valor observado de uma variável aleatória.
- As variáveis são consideradas:
  - independentes e identicamente distribuídas.
- Essas variáveis são conhecidas com uma amostra aleatória.

Exemplo:

- O engenheiro quer medir o enchimento de uma lata de refrigerante.
- Ele retira uma amostra de tamanho n.
- A amostra aleatória é

$$X_1, X_2, \ldots X_n \; .$$

- Não sabemos o valor da primeira medida $X_1 \Rightarrow$ é uma variável aleatória.
- Consideramos que as medidas são independentes e têm a mesma distribuição.

Amostra aleatória

- As variáveis aleatórias

$$X_1, \ldots X_n$$

  são uma amostra aleatória de tamanho n.
- Devem satisfazer as seguintes condições:
  - os X_i's são independentes;
  - cada X_i tem a mesma distribuição de probabilidade.

- Suponha que queremos investigar a proporção de pessoas que preferem uma marca de refrigerante.
- Seja p o valor desconhecido dessa proporção.
- Selecionamos uma amostra de tamanho n.
- A proporção de pessoas $\hat{p}$ que preferem a marca é uma estimativa de p.
- Temos que

$$\hat{p} = \frac{\text{número de indivíduos que preferem a marca}}{n}.$$

- $\hat{p}$ é uma função dos dados observados.
- Dizemos que $\hat{p}$ é uma estatística.

Estatística

- É qualquer função das observações em uma amostra aleatória.

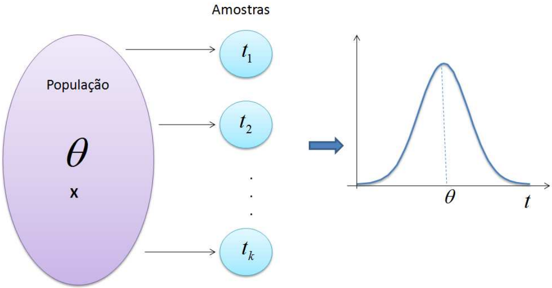
- Exemplos:
  - média amostral
$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$
  - variância amostral
$$S^2 = \frac{\sum_i (X_i - \bar{X})^2}{(n-1)}$$
  - desvio-padrão
$$\sqrt{S^2}\,.$$

- A estatística é uma variável aleatória.
- Ela tem distribuição de probabilidade: distribuição amostral.

Distribuição Amostral

A distribuição de probabilidades de uma estatística é chamada de uma **distribuição amostral**.

- Colhida uma amostra, observamos o valor da estatística T.
- Baseado nesse valor faremos uma afirmação sobre θ.
- Seria interessante saber:
  - como T se comportaria se analisássemos todas amostras possíveis?
  - qual a distribuição de T quando $(X_1, \ldots, X_n)$ assume todos valores possíveis?



Exemplo:

- Lançamos uma moeda honesta três vezes.
- Se sair cara ganhamos 1 ponto, se sair coroa perdemos 1.
- X_i são os pontos obtidos em cada lançamento

$$X_i = \begin{cases} 1 & \text{com probabilidade } 1/2 \\ -1 & \text{com probabilidade } 1/2. \end{cases}$$

- Nesse caso podemos enumerar todas amostras possíveis e calcular $\bar{X}$ e S^2.
- Por exemplo

$$P((X_1, X_2, X_3) = (1, 1, -1)) = \frac{1}{2^3} = \frac{1}{8} \quad \text{e} \quad \bar{X} = 1/3, \quad S^2 = \frac{4}{3}.$$

- Considerando todos os casos encontramos as distribuições de $\bar{X}$ e S^2

$\bar{X}$	-1	-1/3	1/3	1
p_i	1/8	3/8	3/8	1/8

S^2	0	4/3
p_i	1/4	3/4

- Na maioria das vezes não é viável enumerar todos resultados possíveis.
- Precisamos de ferramentas para encontrar as distribuições.

Distribuição da média amostral:

- Uma amostra de tamanho n é retirada de uma população normal.
- Suponha que a média da população é μ e variância σ^2.
- Funções lineares de variáveis normais também são normais.
- Portanto a média:

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

tem distribuição normal com média

$$E(\bar{X}) = \frac{\mu + \cdots + \mu}{n} = \mu$$

e variância

$$Var(\bar{X}) = \frac{\sigma^2 + \cdots + \sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

- Suponha que não conhecemos a distribuição.
- Podemos aproximar a distribuição da média por uma normal:
  - Teorema Central do Limite.

Teorema Central do Limite

- Considere uma amostra de tamanho n

$$X_1, \ldots, X_n .$$

- Essa amostra é retirada de uma população com média μ e variância σ^2.
- Seja $\bar{X}$ a média dessa amostra.
- Então

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

- A aproximação normal para $\bar{X}$ depende do tamanho da amostra n.
- A figura abaixo mostra a distribuição obtida arremessando um dado com seis faces.

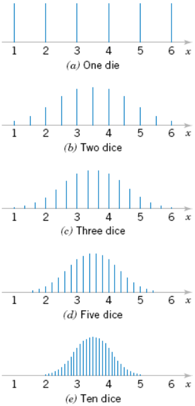


Figure 7-6
Distributions of average scores from throwing dice. [Adapted with permission from Box, Hunter, and Hunter (1978).]

- A distribuição da população está longe da normal.
- Porém as médias são aproximadas razoavelmente por uma normal.
- Geralmente é necessário um tamanho de amostra grande.

- Valores como $n = 4$ ou $n = 5$ não costumam ser suficientes.
- Uma regra prática é usar a aproximação se $n \geq 30$.
- Se $n < 30$ o teorema funcionará se a distribuição da população não for muito diferente da normal.

Exemplo:

- Uma companhia eletrônica fabrica resistores.
- Os resistores têm resistência média de 100 ohms e desvio padrão de 10 ohms.
- A distribuição das resistências é normal.
- Uma amostra aleatória de tamanho $n = 25$ é selecionada.
- Qual probabilidade da média ser menor que 95 ohms?
- Temos que
$$\bar{X} \sim N(\mu, \sigma^2/n) \, .$$
- Portanto a probabilidade requerida é
$$P(\bar{X} < 95) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{95 - 100}{10/\sqrt{25}}\right) = P(Z < -2,5) = 0,0062$$

Exemplo:

- Seja X uma variável com distribuição uniforme

$$f(x) = \begin{cases} \frac{1}{2} & \text{se } 4 \leq x \leq 6 \\ 0 & \text{caso contrário.} \end{cases}$$

- Encontre a distribuição da média de uma amostra aleatória de tamanho $n = 40$.
- Temos que

$$E(X) = \mu = 5 \qquad Var(X) = \sigma^2 = \frac{(6-4)^2}{12} = \frac{1}{3}\,.$$

- Então

$$E(\bar{X}) = \mu = 5 \qquad Var(\bar{X}) = \frac{\sigma^2}{40} = \frac{1/3}{40}\,.$$

- Pelo Teorema do Limite do Central

$$\bar{X} \sim N(5, 1/((3)40))$$

Exemplo: (continuação)

- A figura abaixo mostra os gráficos de X e $\bar{X}$.

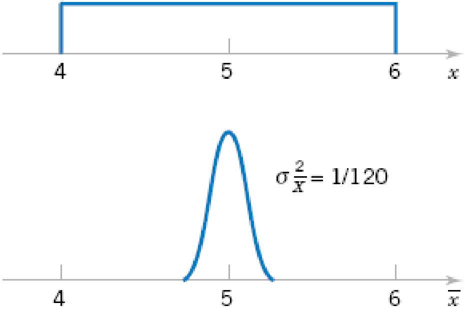


Figure 7-8 The distributions of X and $\bar{X}$ for Example 7-14.

Exemplo:

- Seja X o consumo mensal em minutos por conta de celular de uma região.
- X tem média 40 e desvio padrão 12 minutos.
- Toma-se uma amostra de 24 usuários.
- Encontre:
    - a probabilidade do tempo médio de uso na amostra exceder 45 minutos?
    - a probabilidade do tempo médio de uso na amostra ser menor que 50 minutos?

Exemplo: (solução)

- A média amostral é uma variável com média e variância:

$$E(\bar{X}) = 40 \qquad Var(\bar{X}) = \frac{12^2}{24} = 6\,.$$

- Temos que

$$P(\bar{X} > 45) = P\left(Z > \frac{45-40}{\sqrt{6}}\right) = P(Z > 2,0412) =$$

$$1 - 0,9794 = 0,0206\,.$$

- Além disso

$$P(\bar{X} < 50) = P\left(Z < \frac{50-40}{\sqrt{6}}\right) = P(Z < 4,0824) = 1\,.$$

Exemplo:

- As notas num certo exame padronizado têm média 450 e desvio padrão 50.
- Uma nota acima de 480 é considerada muito boa.
- Uma pessoa entra em uma Universidade se ela obtém acima de 480 neste exame.
- Numa certa sala onde o exame foi aplicado, 25 pessoas fizeram o teste.
- A nota média destas pessoas foi 490.
- Isso é estranho?
- Você acha que houve fraude?
- Dica: use o Teorema Central do Limite.

Exemplo: (solução)

- Seja X a nota no teste.
- X tem média 450 e desvio padrão 50.
- Temos que

$$E(\bar{X}) = 450 \qquad Var(\bar{X}) = \frac{(50)^2}{25} \, .$$

$$P(\bar{X} > 490) = P\left(Z > \frac{490 - 450}{\sqrt{50^2/25}}\right) = P(Z > 4) = 0 \, .$$

- É absolutamente improvável que a nota média daquelas 25 pessoas tenha superado 490.
- Há um indício claro de fraude no teste.

Conceitos Gerais de Estimação pontual

- Os estimadores devem ser escolhidos de forma adequada.
- Devem apresentar determinadas características:
  - não tendencioso;
  - com baixa variância.

Tendenciosidade de um Estimador

- Um estimador $\hat{\theta}$ é não tendencioso para θ se

$$E(\hat{\theta}) = \theta \, .$$

- Se o estimador for tendencioso, a diferença

$$E(\hat{\theta}) - \theta$$

é chamada de tendenciosidade do estimador $\hat{\theta}$.

Estimador Não-tendencioso de Variância Mínima

- Consideramos todos estimados não-tendenciosos de θ.
- Aquele estimador com menor variância:
  - **estimador não-tendencioso de variância mínima**.

Erro Médio Quadrático de um Estimador

- Entre os diversos possíveis estimadores, o EMQ é uma maneira de escolher o melhor estimador.
- Logo o EMQ é um critério de escolha de estimadores.
- Algumas vezes não é óbvio determinar um estimador não tendencioso.
- Em casos mais complexos, estimadore não tendenciosos podem possuir alta variabilide, nesses casos, é possível encontrar estimadores tendenciosos no qual o EMQ seja reduzido.
- Portanto o EMQ é um critério que leva em conta tendência e variabilide.

Erro Médio Quadrático de um Estimador

O **erro médio quadraático** de um estimador $\hat{\theta}$ do parâmetro θ é definido como

$$\text{EMQ}(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = V(\hat{\theta}) + (\text{tendência})^2$$

- Assim, para um estimador não tendencioso ($E(\hat{\theta}) - \theta$), o EMQ é simplesmente $V(\hat{\theta})$.
- O EMQ é um critério importante para escolha de estimadores.
- Seja $\hat{\theta}_1$ e $\hat{\theta}_2$ dois estimadores. O estimador com menor EMQ é preferível.

Exemplo:

- Seja $X_i \sim N(\mu, \sigma^2)$
- Suponha que uma amostra de tamanho 2 seja feita.
- Defina
$$\hat{\theta}_1 = \frac{X_1 + X_2}{2}$$
e
$$\hat{\theta}_2 = \frac{X_1 + 3X_2}{4}$$
- Qual é o melhor estimador?

Exemplo: (solução)

- Ambos estimadores são não tendenciosos

$$E(\hat{\theta}_1) = E(\frac{X_1 + X_2}{2}) = \frac{\mu + \mu}{2} = \mu \, .$$

e

$$E(\hat{\theta}_2) = E(\frac{X_1 + 3X_2}{4}) = \frac{\mu + 3\mu}{4} = \mu \, .$$

- Logo, o EMQ($\hat{\theta}_1$) = $V(\hat{\theta}_1)$ e EMQ($\hat{\theta}_2$) = $V(\hat{\theta}_2)$.
- Assim o melhor estimador é aquele com a menor variância.

- A variância de $\hat{\theta}_1$ é

$$V(\hat{\theta}_1) = V(\frac{X_1 + X_2}{2}) = \frac{\sigma^2 + \sigma^2}{4} = \frac{\sigma^2}{2}\ .$$

- A variância de $\hat{\theta}_2$ é

$$V(\hat{\theta}_2) = V(\frac{X_1 + 3X_2}{4}) = \frac{\sigma^2 + 9\sigma^2}{16} = \frac{5\sigma^2}{8}\ .$$

- Como EMQ($\hat{\theta}_1$) $<$ EMQ($\hat{\theta}_2$), o estimador $\hat{\theta}_1$ é preferível.

Método dos Momentos

- A idéia desse método dos momentos é igualar os momentos populacionais (definidos através da amostra) com os momentos da distribuição.

Momentos

Seja $X_1, \ldots, X_n$ uma amostra de uma população com distribuição de probabilidade $f(x)$. O k-ésimo momento da distribuição é definido por

$$E(X^k),\ k = 1, 2, \ldots\ .$$

O momento k-ésimo populacional correspondente é dado por

$$\frac{1}{n}\sum_{1}^{n} X_i^k,\ k = 1, 2, \ldots\ .$$

- Assim, o método dos momentos é definido igualando-se o momento populacional com o momento da distribuição.
- Para ilustrar, o primeiro momento da distribuição é $E(X^1) = E(X) = \mu$ e o primeiro momento populacional é $\frac{1}{n}\sum_1^n X_i^1 = \bar{X}$.
- Portanto, o estimador de momentos para a média é $\hat{\mu} = \bar{X}$.

Momentos

Seja $X_1, \ldots, X_n$ uma amostra de uma população com distribuição de probabilidade $f(x)$ com m parâmetros desconhecidos $(\theta_1, \ldots, \theta_m)$. Os **estimadores de momento** $\hat{\theta}_1, \ldots, \hat{\theta}_m$ são encontrados igualando os m primeiros momentos da população aos m primeiros momentos da amostra e resolvendos as equações resultantes para os parâmetros desconhecidos.

Exemplo:

- Suponha $X \sim Exp(\lambda)$
- Seja $X_1, \ldots, X_n$ uma amostra coletada dessa distribuição
- Qual é o estimador de momentos de λ?

Exemplo: (solução)

- Como a distribuição só possui um parâmetro desconhecido (λ), devemos igualar o primeiro momento de $f(x)$, $E(X)$, ao primeiro momento populacional $\bar{X}$.
- Dessa forma

$$E(X) = \frac{1}{\lambda} = \bar{X} \rightarrow \hat{\lambda} = \frac{1}{\bar{X}}.$$

Exemplo:

- Suponha $X \sim N(\mu, \sigma^2)$
- Seja $X_1, \ldots, X_n$ uma amostra coletada dessa distribuição
- Qual é o estimador de momentos de $\theta_1 = \mu$ e de $\theta_2 = \sigma^2$?

Exemplo: (solução)

- Como a distribuição possui dois parâmetros desconhecidos (μ e σ^2), devemos igualar os dois primeiros momentos de $f(x)$, $E(X)$ e $E(X^2)$, aos dois primeiro momentos populacionais, $\bar{X}$ e $\frac{1}{n}\sum_1^n X_i^2$.
- Dessa forma

$$E(X) = \mu = \bar{X} \to \hat{\mu} = \bar{X}\,.$$

e

$$E(X^2) = \mu^2 + \sigma^2 = \frac{1}{n}\sum_1^n X_i^2$$

$$\to \hat{\sigma^2} = \frac{\sum_1^n X_i^2 - n(\bar{X})^2}{n} = \frac{\sum_1^n (X_i - \bar{X})^2}{n}\,.$$

Exemplo:

- Suponha $X \sim \text{Gamma}(r, \lambda)$
- Seja $X_1, \ldots, X_n$ uma amostra coletada dessa distribuição
- Qual é o estimador de momentos de $\theta_1 = r$ e de $\theta_2 = \lambda$?

Exemplo: (solução)

- Como a distribuição possui dois parâmetros desconhecidos (r e λ), devemos igualar os dois primeiros momentos de $f(x)$, $E(X)$ e $E(X^2)$, aos dois primeiro momentos populacionais, $\bar{X}$ e $\frac{1}{n}\sum_1^n X_i^2$.
- Lembre-se que $E(X) = \frac{r}{\lambda}$ e $E(X^2) = \frac{r(r+1)}{\lambda^2}$
- Dessa forma resolvendo os sistemas temos

$$\hat{r} = \frac{\bar{X}^2}{(1/n)\sum_1^n X_i^2 - \bar{X}^2}$$

e

$$\hat{\lambda} = \frac{\bar{X}}{(1/n)\sum_1^n X_i^2 - \bar{X}^2}.$$

(fazer no quadro)

Método de Máxima Verossimilhança

- A idéia desse método é, baseando-se nos dados, estimar o(s) parâmetro(s) de tal forma que a probabilidade de ocorrência desses dados observados baseados naquele(s) parâmetro(s) estimado(s) seja máxima.

Exemplo:

- Suponha uma urna com bolas pretas e brancas
- Sabemos que a razão de bolas é de 3 para 1.
- Porém não sabemos para qual cor, ou seja, P(preta)=1/4 ou 3/4.

- Se retiramos $n = 3$ bolas com reposição da urna, e seja X=# de bolas pretas.
- Logo $X \sim Binomial(3, p)$
- Qual é o valor mais plausível para p?

Tabela: Probabilidade Binomial para cada valor possível de p.

x	0	1	2	3
$p(x, 3/4)$	1/64	9/64	27/64	27/64
$p(x, 1/4)$	27/64	27/64	9/64	1/64

- Assim, se para o experimento observamos $n = 2$, qual a proporção de bolas pretas?
- Claramente se $n = 2$ temos que $\hat{p} = 3/4$ é mais provável, dado que $p(2, 3/4) = 27/64 > 9/64 = p(2, 1/4)$.
- Por fim, temos que o estimador de

$$\hat{p} = \begin{cases} 1/4 & \text{se } x = 0, 1 \\ 3/4 & \text{se } x = 2, 3. \end{cases}$$

Função de Verossimilhança

Seja $x_1, \ldots, x_n$ uma amostra observada de uma população com distribuição de probabilidade $f(x)$ com parâmetro θ. A função de verossimilhaça é definida por

$$L(\theta) = f(x_1;\theta) \times \ldots \times f(x_n;\theta) = \prod_{i}^{n} f(x_i;\theta)$$

Note que a função $L(\theta)$ depende somente do(s) parâmetro(s) θ.

- Assim, o **estimador de máximo verossimilhança** de θ é o valor $\hat{\theta}$ que maxima a função $L(\theta)$.

$$\hat{\theta} = argmax_{\theta} L(\theta)$$

- Dessa forma para encontrar $\hat{\theta}$ devemos maximizar $L(\theta)$.

Exemplo:

- Suponha $X \sim N(\theta, \sigma^2)$
- Seja $X_1, \ldots, X_n$ uma amostra coletada dessa distribuição
- Qual é o estimador de máxima verossimilhança de θ?

Exemplo: (solução)

- Como os dados vêm de uma distribuição normal, temos que

$$
\begin{aligned}
L(\theta) &= \prod_1^n f(x_i) = \prod_1^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\theta)^2/2\sigma^2} \\
&= \frac{1}{(2\pi\sigma^2)^2} e^{-1/2\sigma^2 \sum_1^n (x_i-\theta)^2}
\end{aligned}
$$

- Assim, para maximizar uma função devemos deriva-la e igualar a 0. Antes iremos calcular o $\log(L(\theta)) = l(\theta)$,

$$
l(\theta) = -(n/2)\log(2\pi\sigma^2) - (2\sigma^2)^{-1} \sum_1^n (x_i - \theta)^2
$$

- Derivando $l(\theta)$ temos

$$\frac{dl(\theta)}{d\theta} = -(2\sigma^2)^{-1}\sum_{1}^{n}(x_i - \theta)^2 \equiv 0.$$

- Resolvendo temos

$$\hat{\theta} = 1/n\sum_{i}^{n} X_i = \bar{X}$$

(fazer no quadro)

Exemplo:

- Suponha $X \sim Exp(\theta)$
- Seja $X_1, \ldots, X_n$ uma amostra coletada dessa distribuição
- Qual é o estimador de máxima verossimilhança de θ?

Exemplo: (solução)

- Como os dados vêm de uma distribuição exponencial, temos que

$$L(\theta) = \prod_{1}^{n} f(x_i) = \prod_{1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_i^n x_i}$$

- Assim, para maximizar uma função devemos deriva-la e igualar a 0. Antes iremos calcular o $\log(L(\theta)) = l(\theta)$,

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

- Derivando $l(\theta)$ temos

$$\frac{dl(\theta)}{d\theta} = \frac{n}{\theta} - \sum_{i}^{n} x_i \equiv 0.$$

- Resolvendo temos

$$\hat{\theta} = n/\sum_{i}^{n} X_i = 1/\bar{X}$$

(fazer no quadro)