Capítulo 6 - Amostragem e Estimação Pontual

Conceição Amado, Ana M. Pires e M. Rosário Oliveira

6.1 Inferência Estatística. Amostragem aleatória

Definição: Uma população é o conjunto de todas as observações possíveis de determinada variável de interesse, X.

$$\text{Quanto ao tamanho} \begin{cases} \text{finitas} & \begin{cases} \text{pequenas} \\ \text{grandes} \end{cases} \\ & \quad \swarrow\approx \\ \text{infinitas} & \end{cases}$$

Exemplos:

- Alturas da população portuguesa
- Idades dos estudantes inscritos nesta turma
- Resistências dos filamentos de um conjunto de lâmpadas
- Temperaturas em todos os pontos de uma sala

Por conveniência identificamos a população com a v.a. correspondente, X.

6.1 (cont.)

Para conhecer exactamente X teríamos de fazer um número muito grande ou infinito de observações. Isso pode ser impossível, muito caro, ou muito demorado.

Podemos obter algum conhecimento sobre X se observarmos alguns valores da população $\longrightarrow$ AMOSTRA.

No entanto, a inferência das propriedades da população só é possível se a amostra for obtida por métodos probabilísticos (ou seja, se for fixada a probabilidade de um qualquer elemento da população vir a pertencer à amostra).

Amostragem (probabilística)

$f_X(x) = ?$

$\mu_X = ?$

$\sigma_X = ?$

$\dots$

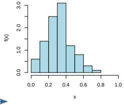
POPULAÇÃO

X

Amostra

$x_1, x_2, \dots, x_n$

Estatística descritiva



$\bar{x} = 0.32$

$s = 0.14$

$\dots$

Inferência estatística

(ou estatística indutiva)

6.1 (cont.)

Conhecer exactamente X corresponde a conhecer a sua função de distribuição $F_X(x)$

(o que é equivalente a conhecer a f.d.p. caso X seja contínua, ou a f.p. caso X seja discreta)

Admitem-se dois níveis de "ignorância":

1. $F_X(x)$ é completamente desconhecida, sabendo-se apenas se é do tipo contínuo ou discreto;
2. Admite-se (pelo conhecimento do fenómeno em causa) que $F_X(x)$ pertence a determinada família, por exemplo normal ou Poisson, mas com parâmetros (nesse exemplo μ e σ, ou λ) desconhecidos.

6.1 (cont.)

Objectivos da Inferência Estatística:

- estimar $F_X(x)$ ou estimar os parâmetros de $F_X(x)$ conhecendo a sua forma;
- fazer testes em relação aos parâmetros ou em relação à forma de $F_X(x)$.

Estimação de parâmetros:

- pontualmente $\rightarrow$ resto do Capítulo 6;
- por intervalo $\rightarrow$ Capítulo 7.

Testes de hipóteses $\rightarrow$ Capítulo 8:

- sobre parâmetros $\rightarrow$ Capítulo 8;
- sobre a forma de $F_X(x)$ $\rightarrow$ Capítulo 8.

Amostragem aleatória

O processo de amostragem probabilística que vamos considerar pode ser descrito informalmente do seguinte modo: cada elemento da amostra é obtido totalmente ao acaso na população e de forma independente dos outros elementos.

Desta forma cada elemento da amostra é o valor observado de uma variável aleatória com distribuição idêntica à população e essas variáveis aleatórias são independentes.

Definição: As variáveis aleatórias $(X_1, X_2, \ldots, X_n)$ constituem uma **amostra aleatória** de dimensão n da população X se forem **i**ndependentes e **i**denticamente **d**istribuídas a X (i.i.d.)

$$\begin{array}{ccccccll} X & \longrightarrow & X_1 & X_2 & \cdots & X_n & n \text{ v.a. i.i.d a X} & \textbf{amostra aleatória} \\ & & \downarrow & \downarrow & & \downarrow & & \\ & & x_1 & x_2 & \cdots & x_n & n \text{ observações} & \textbf{amostra casual} \end{array}$$

Amostragem aleatória: exemplo

Exemplo 6.1: Seja X a população que corresponde ao n.ºde caras observado no lançamento de uma moeda não necessariamente perfeita. O modelo é conhecido, $X \sim Bernoulli(p)$ $(0 < p < 1)$. O objectivo da amostragem será obter informação sobre p

$X = 1$		cara	$P(X = 1) = p$
$X = 0$		coroa	$P(X = 0) = 1 - p$

8 amostras de dimensão 10:

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
0	1	1	0	1	1	0	0	0	1
1	1	1	1	1	0	1	1	1	1
0	1	0	1	0	1	0	0	0	1
0	0	0	1	1	1	1	0	1	1
1	1	0	0	0	1	1	0	0	0
1	0	1	0	0	0	0	0	0	0
1	1	0	1	0	1	0	0	1	0
1	0	1	0	0	0	0	1	1	1
$P(X_1 = 1) = p$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$P(X_{10} = 1) = p$

Amostragem aleatória: exemplo (cont.)

Na realidade dispomos apenas de uma amostra casual, por exemplo a primeira:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0	1	1	0	1	1	0	0	0	1

mas para poder inferir correctamente a partir desta precisamos de saber que ela é uma de muitas possíveis que se distribuem de determinado modo.

Em particular podemos calcular a probabilidade de observar aquela particular amostra:

$P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0, \ldots, X_7 = 0, X_8 = 0, X_9 = 0, X_{10} = 1) =$

(porque as observações são independentes)

$= P(X_1 = 0)P(X_2 = 1)P(X_3 = 1) \cdots P(X_8 = 0)P(X_9 = 0)P(X_{10} = 1) =$

(porque as observações são identicamente distribuídas a X)

$= (1-p) \times p \times p \times (1-p) \times p \times p \times (1-p) \times (1-p) \times (1-p) \times p = p^5(1-p)^5$

6.1 (cont.)

SUMÁRIO:

- $(X_1, X_2, \ldots, X_n)$ – **amostra aleatória**
  (v.a. de dimensão n que pretende representar "todas" as possíveis amostras dessa dimensão)

- $$\left.\begin{array}{l} (x_1,\, x_2, \ldots, x_n) \\ (x_1,\, x_2,\, x_3,\, x_4,\, x_5) \\ (0,\, 0,\, 1,\, 0,\, 1,\, 0,\, 0) \\ (2,\, 1.5,\, 3,\, 4.2,\, 0.1) \end{array}\right\} \textbf{amostras casuais } \text{(ou concretas)}$$

- $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$

  (probabilidade, ou densidade de probabilidade, de observar a amostra $(x_1, x_2, \ldots, x_n)$, para uma população com função de probabilidade, ou densidade de probabilidade, $f_X(\cdot)$)

Estatísticas

Em geral usamos funções da amostra para "estimar" certos aspectos da população: por exemplo, é intuitivo perceber que a média da amostra será uma "aproximação" ou "estimativa" possível da média da população.

Definição: Uma estatística é uma v.a. que é função unicamente da amostra aleatória. Denota-se usualmente por $T_n = T(X_1, X_2, \cdots, X_n)$

Exemplos de estatísticas:

1) Média amostral $$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

É uma variável aleatória!

Dada uma amostra concreta $(x_1, x_2, \ldots, x_n)$, podemos calcular o valor da sua média $\bar{x} = (x_1 + x_2 + \cdots + x_n)/n$ que será um valor observado (ou uma ocorrência, ou ainda uma concretização) da v.a. $\bar{X}$

6.1 (cont.)

2) Variância amostral

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$ É uma variável aleatória!

A variância de uma amostra concreta

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

é um valor observado da v.a. S^2.

3) Mínimo da a.a.: $X_{(1)} = \min\{(X_1, X_2, \cdots, X_n\}$ É uma variável aleatória!

4) Máximo da a.a.: $X_{(n)} = \max(X_1, X_2, \cdots, X_n)$ É uma variável aleatória!

6.2 Estimação pontual. Propriedades dos estimadores

Como as estatísticas são variáveis aleatórias faz sentido falar da sua distribuição de probabilidades (que se chama de **distribuição amostral** ou **distribuição por amostragem**).

1. Distribuição amostral do Máximo da a.a. ($X_{(n)}$)

$$F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = P(X_1 \leq x, X_2 \leq x, \cdots, X_n \leq x) =$$

$$= \prod_{i=1}^{n} P(X_i \leq x) = \prod_{i=1}^{n} P(X \leq x) = (F_X(x))^n$$

2. (Fazer para o mínimo.)

Veremos mais adiante as distribuições amostrais de $\bar{X}$ e S^2.

6.2 Estimação pontual. Propriedades dos estimadores

Definição: Chama-se **estimador** a qualquer estatística, $\hat{\Theta}$, usada para estimar um parâmetro, θ (desconhecido) da população ou uma função desse parâmetro. A um valor desse estimador, $\hat{\theta}$, chama-se estimativa.

Seja X – população com $f_X(x)$ que depende de um parâmetro desconhecido θ (pode-se generalizar para vectores de parâmetros).

$(X_1, X_2, \ldots, X_n)$ – a.a.

$\hat{\Theta} = T_n(X_1, X_2, \ldots, X_n)$: estimador pontual de θ

$\hat{\theta} = t_n(x_1, x_2, \ldots, x_n)$: estimativa pontual de θ

Exemplo 6.2: $X \sim Bernoulli(p)$, dada uma a.a. de dimensão n considerar o estimador

$$T_n = \hat{P} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Como $X_i = 1$ se ocorrer um sucesso e $X_i = 0$ se ocorrer um insucesso, conclui-se que $\sum_{i=1}^{n} X_i$ representa o número de sucessos na amostra aleatória (observar que o número de insucessos será $n - \sum_{i=1}^{n} X_i$).

6.2 (cont.)

Exemplo 6.2 (cont.): E $\sum_{i=1}^{n} x_i$ é o número de sucessos na amostra concreta. Por exemplo para a primeira amostra do lançamento da moeda,

0 1 1 0 1 1 0 0 0 1

tem-se $\hat{p} = \bar{x} = 5/10 = 0.5$, que é uma estimativa do parâmetro p.

Definição: O estimador pontual $\hat{\Theta}$ é um estimador centrado do parâmetro θ se $E(\hat{\Theta}) = \theta$.

Se o estimador não for centrado (também se diz, se o estimador for enviesado) chama-se enviesamento (ou viés) à diferença

$$b(\hat{\Theta}) = E(\hat{\Theta}) - \theta$$

6.2 (cont.)

Exemplo 6.3: Dada uma v.a. X com valor esperado μ e variância σ^2 (e distribuição qualquer), tem-se que $\bar{X}$ e S^2 são estimadores centrados de μ e σ^2, respectivamente.

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{E(X_1) + E(X_2) + \cdots + E(X_n)}{n} = \\ &= \frac{\mu + \mu + \cdots + \mu}{n} = \frac{n\mu}{n} = \mu \end{aligned}$$

logo $\bar{X}$ é estimador centrado de μ.
Calculemos também a variância de $\bar{X}$:

$$\begin{aligned} V(\bar{X}) &= V\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{V(X_1) + V(X_2) + \cdots + V(X_n)}{n^2} = \\ &= \frac{\sigma^2 + \sigma^2 + \cdots + \sigma^2}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

6.2 (cont.)

Quanto ao estimador S^2, iniciemos por reescrevê-lo:

$$\begin{aligned} S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^2-2\bar{X}X_i+\bar{X}^2) = \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-2\bar{X}\sum_{i=1}^{n}X_i+n\bar{X}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right) \end{aligned}$$

pois $\sum_{i=1}^{n} X_i = n\bar{X}$. Vamos precisar de calcular $E(X_i^2)$ e $E(\bar{X}^2)$:

$$E(X_i^2) = V(X_i) + E^2(X_i) = \sigma^2 + \mu^2 \qquad E(\bar{X}^2) = V(\bar{X}) + E^2(\bar{X}) = \frac{\sigma^2}{n} + \mu^2$$

logo $E(S^2) = \dfrac{1}{n-1}\left[n(\sigma^2+\mu^2) - n\left(\dfrac{\sigma^2}{n}+\mu^2\right)\right] = \dfrac{1}{n-1}(n-1)\sigma^2 = \sigma^2$

e conclui-se que S^2 é estimador centrado de σ^2

6.2 (cont.)

Exemplo: Se $X \sim Ber.(p)$ sabe-se que $E(X) = \mu = p$ e $V(X) = \sigma^2 = p(1-p)$. Dada uma a.a. de dimensão n e o estimador $\hat{P} = \bar{X} = \sum_{i=1}^{n} X_i/n$, os resultados obtidos para a média amostral permitem afirmar que

$$E(\hat{P}) = p \qquad \text{e} \qquad V(\hat{P}) = \sigma^2_{\hat{P}} = \frac{p(1-p)}{n}$$

ou seja, $\hat{P}$ é um estimador centrado de p e o respectivo desvio padrão (que é uma medida do erro associado à estimativa, também chamada **erro padrão**) é $\sigma_{\hat{P}} = \sqrt{p(1-p)/n}$. Uma estimativa do erro padrão é $\hat{\sigma}_{\hat{P}} = \sqrt{\hat{p}(1-\hat{p})/n}$.

Para a primeira amostra do lançamento da moeda,

● 0 ● 1 ● 1 ● 0 ● 1 ● 1 ● 0 ● 0 ● 0 ● 1

tem-se $\hat{p} = \bar{x} = 5/10 = 0.5$ (como se viu) e $\hat{\sigma}_{\hat{P}} = \sqrt{0.5 \times 0.5/10} \simeq 0.158$.

Notar que para o mesmo $\hat{p} = 0.5$ mas com $n = 1000$ vinha $\hat{\sigma}_{\hat{P}} \simeq 0.0158$

6.2 (cont.)

Como para um mesmo parâmetro pode haver vários estimadores centrados são necessários outros critérios para comparar estimadores.

Definição: O erro quadrático médio de um estimador $\hat{\Theta}$ do parâmetro θ é

$$MSE(\hat{\Theta}) \equiv EQM(\hat{\Theta}) = E\left(\hat{\Theta} - \theta\right)^2$$

Nota: $EQM(\hat{\Theta}) = V(\hat{\Theta}) + b^2(\hat{\Theta})$

Definição: Dados dois estimadores $\hat{\Theta}_1$ e $\hat{\Theta}_2$ de um mesmo parâmetro θ, diz-se que $\hat{\Theta}_1$ é mais eficiente que $\hat{\Theta}_2$ se $MSE(\hat{\Theta}_1) < MSE(\hat{\Theta}_2)$ Ao quociente $MSE(\hat{\Theta}_1)/MSE(\hat{\Theta}_2)$ chama-se eficiência relativa de $\hat{\Theta}_2$ em relação a $\hat{\Theta}_1$.

Dados dois estimadores deve preferir-se o que for mais eficiente, ou seja, o que tiver menor erro quadrático médio.

6.2 (cont.)

Exemplo 6.1 (cont.): Vamos considerar novamente o exemplo da moeda ($X \sim Ber.(p)$ com $0 < p < 1$ desconhecido) e dois estimadores, o que já apareceu antes, agora designado $\hat{P}_1$, e um estimador alternativo $\hat{P}_2$, chamado estimador de Laplace:

$$\hat{P}_1 = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \hat{P}_2 = \frac{1 + \sum_{i=1}^{n} X_i}{n+2}$$

Vamos ver o que acontece com as 8 amostras que foram geradas anteriormente.

6.2 (cont.)

8 amostras de dimensão 10:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	$\hat{p}_1$	$\hat{p}_2$
0	1	1	0	1	1	0	0	0	1	0.5	0.5
1	1	1	1	1	0	1	1	1	1	0.9	0.8333
0	1	0	1	0	1	0	0	0	1	0.4	0.4167
0	0	0	1	1	1	1	0	1	1	0.6	0.5833
1	1	0	0	0	1	1	0	0	0	0.4	0.4167
1	0	1	0	0	0	0	0	0	0	0.2	0.25
1	1	0	1	0	1	0	0	1	0	0.5	0.5
1	0	1	0	0	0	0	1	1	1	0.5	0.5
média dos 8 valores:										0.5	0.5
variância dos 8 valores:										0.04	0.0278

Sabendo que os dados foram gerados com $p = 0.5$, conclui-se que o segundo estimador é melhor (as estimativas estão a distância menor ou igual do verdadeiro valor).

6.2 (cont.)

Sabemos já que
$$EQM(\hat{P}_1) = p(1-p)/n$$
pois $\hat{P}_1$ é centrado ($b(\hat{P}_1) = 0$), logo $EQM(\hat{P}_1) = V(\hat{P}_1)$.

Pode mostrar-se (**Exercício**) que

$$EQM(\hat{P}_2) = \frac{np(1-p) + (1-2p)^2}{(n+2)^2}$$

e que $EQM(\hat{P}_2)$ é mais eficiente que $EQM(\hat{P}_1)$ se

$$p \in \left] \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{n}{8n+4}}\ ;\ \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{n}{8n+4}} \right[$$

6.3 Método da máxima verosimilhança

Existem vários métodos de estimação de parâmetros desconhecidos. Um desses métodos é o da máxima verosimilhança. Como o seu nome indica o estimador obtém-se maximizando uma certa função chamada função de verosimilhança.

Definição: Seja X uma v.a. com distribuição caracterizada por $f(x,\theta)$ (f.p. ou f.d.p.), onde θ é um parâmetro desconhecido. Sejam $x_1, x_2, \dots, x_n$ os valores observados de uma a.a. de dimensão n. A função de verosimilhança da amostra é

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

Chama-se estimativa de máxima verosimilhança de θ ($\hat{\theta}$) ao valor de θ que maximiza $L(\theta)$, ou seja,

$$\hat{\theta} = \arg\max_{\theta} L(\theta; x_1, x_2, \dots, x_n)$$

6.3 (cont.)

Exemplo 1: No exemplo concreto que temos vindo a considerar, $X \sim Bernoulli(p)$, tem-se, se considerarmos a amostra 1,

$$L(p) = p^5(1-p)^5, \qquad 0 < p < 1$$

Determinação do valor de p que maximiza $L(p)$:

$$\frac{dL(p)}{dp} = 5p^4(1-p)^5 - 5p^5(1-p)^4 = 5p^4(1-p)^4(1-2p) = 0$$

$$\Leftrightarrow \quad p = 0 \ \vee \ p = 1 \ \vee \ p = \frac{1}{2}$$

	0		1/2		1
$L'(p)$	0	$+$	0	$-$	9
$L(p)$	0	$\nearrow$	$+$	$\searrow$	0

Logo a estimativa de m.v. de p com base nesta amostra é $\hat{p} = \frac{1}{2}$

6.3 (cont.)

Em vez de fazer a determinação da estimativa para uma amostra concreta é conveniente fazê-lo para uma amostra genérica $(x_1, \ldots, x_n)$

(vantagens: só é preciso fazer os cálculos uma vez e obtém-se a expressão do estimador, necessária para estudar as suas propriedades).

Exemplo 2: $X \sim Ber.(p)$, para a qual $f(x) = P(X = x) = p^x(1-p)^{1-x}$, $0 < p < 1$. Dada uma amostra $(x_1, \ldots, x_n)$ vem

$$\begin{aligned} L(p; x_1, \ldots, x_n) \equiv L(p) &= f(x_1)\cdots f(x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = \\ &= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} = \\ &= p^k(1-p)^{n-k}, \qquad 0 < p < 1 \end{aligned}$$

onde $k = \sum_{i=1}^n x_i$ é o número de sucessos na amostra e $n-k$ o número de insucessos.

6.3 (cont.)

Nota: em vez de determinar p que maximiza $L(p)$ pode determinar-se o valor de p que maximiza $\log L(p)$ (pois para uma função $f > 0$ qualquer, f e $\log f$ têm máximo e mínimo nos mesmos pontos e os cálculos com $\log L$ são geralmente mais simples do que com L).

$$\log L(p) = \log\left[p^k(1-p)^{n-k}\right] = k\log p + (n-k)\log(1-p)$$

$$\frac{d\log L(p)}{dp} = \frac{k}{p} - \frac{n-k}{1-p} = 0 \overset{(p\neq 0,\, p\neq 1)}{\Leftrightarrow} k(1-p) - (n-k)p = 0 \Leftrightarrow p = \frac{k}{n}$$

Verificação: $\displaystyle\frac{d^2\log L(p)}{dp^2} = -\frac{k}{p^2} - \frac{n-k}{(1-p)^2} < 0,\ \forall_{0<p<1}$

a estimativa de m.v. é $\displaystyle\hat{p} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$

o estimador de m.v. é $\displaystyle\hat{p} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$

6.3 (cont.)

O método da máxima verosimilhança também pode ser usado quando a função de densidade (ou de probabilidade) da população depende de mais de um parâmetro.

Exemplo 3: Considere-se $X \sim N(\mu, \theta)$, $(\theta = \sigma^2 > 0)$ $f(x) = \dfrac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-\mu)^2}{2\theta}}$.

Dada uma amostra $(x_1, \ldots, x_n)$ vem

$$\begin{aligned} L(\mu, \theta; x_1, \ldots, x_n) &\equiv L(\mu, \theta) = f(x_1) \cdots f(x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x_i-\mu)^2}{2\theta}} = \\ &= \frac{1}{(2\pi\theta)^{n/2}} e^{-\frac{1}{2\theta} \sum_{i=1}^{n} (x_i-\mu)^2},\ \mu \in \mathbb{R},\ \theta > 0 \end{aligned}$$

$$\log L(\mu, \theta) = -\frac{n}{2} \log(2\pi\theta) - \frac{1}{2\theta} \sum_{i=1}^{n} (x_i - \mu)^2$$

6.3 (cont.)

Determinação do ponto de máximo:

$$\begin{cases} \dfrac{\partial \log L(\mu,\theta)}{\partial \mu} = 0 \\[2ex] \dfrac{\partial \log L(\mu,\theta)}{\partial \theta} = 0 \end{cases} \Leftrightarrow \begin{cases} \dfrac{1}{\theta}\displaystyle\sum_{i=1}^{n}(x_i - \mu) = 0 \\[2ex] -\dfrac{n}{2\theta} + \dfrac{1}{2\theta^2}\displaystyle\sum_{i=1}^{n}(x_i - \mu)^2 = 0 \end{cases}$$

$$\begin{cases} \hat{\mu} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i \\[2ex] \hat{\theta} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 \end{cases}$$

solução candidata
é necessário confirmar
que corresponde de facto
a um ponto de máximo

6.3 (cont.)

Para isso deve-se analisar o determinante e a diagonal da matriz Hessiana no ponto $(\mu, \theta) = (\hat{\mu}, \hat{\theta})$

$$\left[\begin{array}{cc} \dfrac{\partial^2 \log L(\mu,\theta)}{\partial \mu^2} & \dfrac{\partial^2 \log L(\mu,\theta)}{\partial \mu \partial \theta} \\ \dfrac{\partial^2 \log L(\mu,\theta)}{\partial \mu \partial \theta} & \dfrac{\partial^2 \log L(\mu,\theta)}{\partial \theta^2} \end{array}\right]_{(\mu,\theta)=(\hat{\mu},\hat{\theta})} = \left[\begin{array}{cc} -\dfrac{n}{\hat{\theta}} & 0 \\ 0 & -\dfrac{n}{2\hat{\theta}^2} \end{array}\right]$$

Como o determinante é positivo e os elementos da diagonal principal são ambos negativos está confirmado que a solução encontrada corresponde de facto a um ponto de máximo da função $L(\mu, \theta)$.

Logo **os estimadores de m.v.** de μ e $\theta = \sigma^2$ são

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

(o primeiro é centrado e o segundo não)

6.3 (cont.)

Nota: Os estimadores de máxima verosimilhança não são necessariamente centrados mas são assintoticamente centrados (quando $n \to \infty$)

Propriedade da invariância dos estimadores de máxima verosimilhança:

Se $\hat{\Theta}_1$, $\hat{\Theta}_2$, ..., $\hat{\Theta}_k$, são estimadores de máxima verosimilhança dos parâmetros θ_1, θ_2, ..., θ_k, então o estimador de máxima verosimilhança de uma função $h(\theta_1, \theta_2, \ldots, \theta_k)$ desses parâmetros é a mesma função $h(\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k)$ dos estimadores.

Exemplo Seja $X_1, X_2, \cdots, X_n$ uma a.a. proveniente de uma população X com distribuição Poisson de parâmetro λ.

Determine os estimadores de máxima verosimilhança de λ e de $P(X > 2)$.

6.4 Momentos da média amostral e de variâncias amostrais. Distribuições amostrais da média ...

Momentos:

- O momento de ordem k de uma v.a. X é $E(X^k)$.

  O primeiro momento é o valor esperado, $E(X) = \mu$

- O momento central de ordem k de uma v.a. X é $E[(X-\mu)^k]$.

  O segundo momento central é a variância, $E[(X-\mu)^2] = V(X) = \sigma^2$

Os momentos das estatísticas média e variância amostrais foram já calculados:

$$\boxed{E(\bar{X}) = \mu} \qquad \boxed{V(\bar{X}) = \frac{\sigma^2}{n}} \qquad \boxed{E(S^2) = \sigma^2}$$

Estes são os únicos momentos da média e da variância amostrais que não dependem da distribuição da população.

6.4 ... Distribuições amostrais da média e variância numa população normal ...

A distribuição de probabilidades de uma estatística é chamada distribuição amostral ou distribuição por amostragem.

Teorema Considere-se uma população $X \sim N(\mu, \sigma^2)$ e uma a.a. $(X_1, X_2, \ldots, X_n)$. Como

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

é uma combinação linear de variáveis aleatórias independentes com distribuição normal, conclui-se que também tem distribuição normal, logo

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \Leftrightarrow \qquad \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

6.4 ... Distribuições amostrais da média e variância numa população normal ...

Para populações não normais tem-se **como consequência do T.L.C.:**

Se $(X_1, X_2, \ldots, X_n)$ for uma amostra aleatória de dimensão n de uma população X com valor esperado μ e variância σ^2 e $\bar{X}$ a correspondente média amostral então

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{a}{\sim} N(0, 1)$$