Distribuições Amostrais

Estimador é uma variável aleatória e portanto, possui uma distribuição de probabilidades, denominada distribuição amostral do estimador.

No exemplo

X	0	10	20	30
p_i	0,2	0,3	0,3	0,2

$$\hat{\mu} = \frac{X_1 + X_2}{2}$$ estimador de $\mu = E(X)$ $\hat{\mu} = \bar{X}$

amostra aleatória $n = 2$

Distribuição Amostral de $\bar{X}$

$\bar{X}$	0	5	10	15	20	25	30
p_i	0,04	0,12	0,21	0,26	0,21	0,12	0,04

Da mesma maneira, poderia ser obtida a distribuição amostral de S^2.

Continuaremos a estudar a distribuição amostral de $\bar{X}$ e os resultados serão exatos ou aproximados, dependendo da distribuição da variável original X.

Resultado

Variável de interesse $X \sim N(\mu, \sigma^2)$

$X_1, X_2 \dots X_n$ amostra aleatória dessa densidade:

$X_i \sim N(\mu, \sigma^2)$ $i = 1, 2 \dots n$

X_i, X_j são independentes para $i \neq j$.

Nessas condições,

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

Prova:

Do tópico anterior

$X_1, X_2 \dots X_n$ v.a. independentes $\sim N(\mu_i, \sigma_i^2)$

$$\Rightarrow \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum a_i^2 \sigma_i^2\right)$$

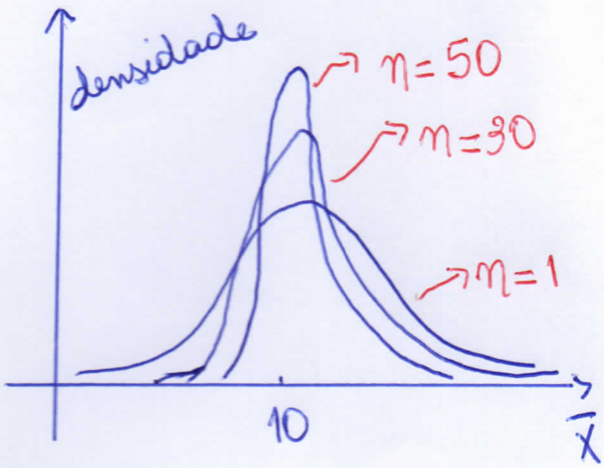
No caso, $\bar{X} = \sum a_i X_i$ para $a_i = \frac{1}{n}$ $i = 1, 2 \dots n$

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} \frac{1}{n}\mu = n\frac{\mu}{n} = \mu$$

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum a_i^2 \sigma_i^2 = \sum_{i=1}^{n} \frac{1}{n^2}\sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Consequência:

$X \sim N(10, 16) \Rightarrow \bar{X} \sim N(10, 16/n)$ → pag 237



À medida que $\underline{n}$ aumenta, a densidade de $\bar{X}$ vai se concentrando em torno de $\mu = 10$.

Quanto maior n, maior a probab. de $\bar{X}$ próximo da média populacional μ.

Aplicações do Resultado

Exemplo 1 (exemplo 7.14, Magalhães e Lima, pag 238)

Um lote de peças é aceito se o comprimento médio de uma amostra de 10 peças, retiradas aleatoriamente do lote, estiver entre 5 e 10 cm.

Informação : Comprimento $X \sim N(7,5 ; 20)$

Qual é a probabilidade de aceitação do lote?

X_i - comprimento da i-ésima peça da amostra

$i = 1, 2, \ldots 10$

$X_i \sim N(7,5 ; 20)$ amostra aleatória $\Rightarrow \bar{X} \sim N\left(7,5 ; \frac{20}{10}\right)$

$\frac{\sigma^2}{n}$

$$P(\text{aceitação}) = P(5 < \bar{X} < 10) =$$

$$= P\left(\frac{5-7,5}{\sqrt{2}} < \frac{\bar{X}-7,5}{\sqrt{2}} < \frac{10-7,5}{\sqrt{2}}\right) = P(-1,77 < Z < 1,77)$$

$Z \sim N(0,1)$

$$= 2 \times P(0 < Z < 1,77) = 0,9232$$

Exemplo 2

Amostra de tamanho 25 de uma distribuição normal com média μ e variância σ^2.

Calcular $P(|\bar{X}-\mu| \leq 1)$ para $\sigma^2 = 16, 64$ e 100.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{25}\right) \qquad Z = \frac{\bar{X}-\mu}{\sigma/5} \sim N(0,1)$$

a) $\sigma = 4$

$$P(|\bar{X}-\mu| \leq 1) = P(-1 \leq \bar{X}-\mu \leq 1) = P\left(\frac{-1}{4/5} \leq \frac{\bar{X}-\mu}{4/5} \leq \frac{1}{4/5}\right)$$

$$= P(-1{,}25 \leq Z \leq 1{,}25) = 2 \times 0{,}3944 = 0{,}7888$$

A probabilidade do estimador $\bar{X}$ (da média μ) não se distanciar de μ por mais que 1 unidade é 0,7888

b) $\sigma = 8$

$$P(|\bar{X}-\mu| \leq 1) = P(-1 \leq \bar{X}-\mu \leq 1) = P\left(\frac{-1}{8/5} \leq \frac{\bar{X}-\mu}{8/5} \leq \frac{1}{8/5}\right)$$

$$= P(-0{,}625 \leq Z \leq 0{,}625) = 2 \times 0{,}2357 = 0{,}4714$$

c) $\sigma = 10$

$$P\left(\frac{-1}{10/5} \leq \frac{\bar{X} - \mu}{10/5} \leq \frac{1}{10/5} \right) = P(-0{,}5 \leq Z \leq 0{,}5) =$$

$$= 2 \times 0{,}1915 = 0{,}3830.$$

d) O que aconteceria se repetíssemos para n maior, por ex., $n = 81$?

$\sigma = 4$ $\quad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \bar{X} \sim N\left(\mu, \frac{16}{81}\right)$

$$Z = \frac{\bar{X} - \mu}{4/9} \sim N(0,1)$$

$$P(-1 \leq \bar{X} - \mu \leq 1) = P\left(\frac{-9}{4} \leq \frac{\bar{X} - \mu}{4/9} \leq \frac{9}{4} \right) =$$

$$= P(-2{,}25 \leq Z \leq 2{,}25) = 2 \times 0{,}4878 = 0{,}9756$$

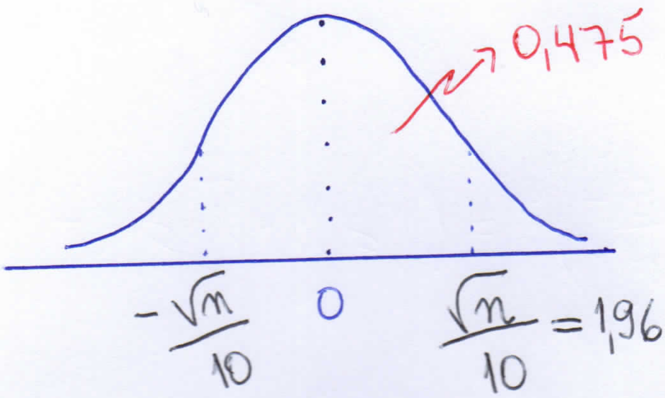
(Para $n = 25$ era $0{,}7888$)

e) Qual é o tamanho de amostra necessário para que $P(|\bar{X}-\mu| \leq 1) = 0{,}95$ se $\sigma^2 = 100$?

$$\bar{X} \sim N\left(\mu, \frac{100}{n}\right)$$

$$P(-1 \leq \bar{X}-\mu \leq 1) = P\left(\frac{-1}{10/\sqrt{n}} \leq \frac{\bar{X}-\mu}{10/\sqrt{n}} \leq \frac{1}{10/\sqrt{n}}\right) =$$

$$= P\left(-\frac{\sqrt{n}}{10} \leq Z \leq \frac{\sqrt{n}}{10}\right) = 0{,}95 \qquad Z \sim N(0,1)$$



$$\frac{\sqrt{n}}{10} = 1{,}96$$

$$\sqrt{n} = 19{,}6$$

$$n = 384{,}16 \simeq 385$$

e para $P(|\bar{X}-\mu| \leq 0{,}5) = 0{,}95$?

$$P(-0{,}5 \leq \bar{X}-\mu \leq 0{,}5) = P\left(-\frac{0{,}5}{10/\sqrt{n}} \leq \frac{\bar{X}-\mu}{10/\sqrt{n}} \leq \frac{0{,}5}{10/\sqrt{n}}\right)$$

$$= P(-0{,}05\sqrt{n} \leq Z \leq 0{,}05\sqrt{n}) = 0{,}95$$

$$0{,}05\sqrt{n} = 1{,}96 \Rightarrow n = 1536{,}64 \simeq 1537$$

$P(|\bar{X} - \mu \leq 1)$

	$n = 25$	81	385	1537
$\sigma = 4$	0,7888	0,9756		
8	0,4714			
10	0,3830		0,95	0,95

diminui

aumenta

Completar a tabela

Se X não tem distribuição Normal utiliza-se

Teorema do Limite Central

Se $X_1, X_2 \ldots X_n$ é uma amostra aleatória de tamanho n da v.a. X com média μ e Variância σ^2 então

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \xrightarrow[n \to \infty]{} N(0,1)$$

(Converge em distribuição para uma v.a. com distribuição $N(0,1)$)

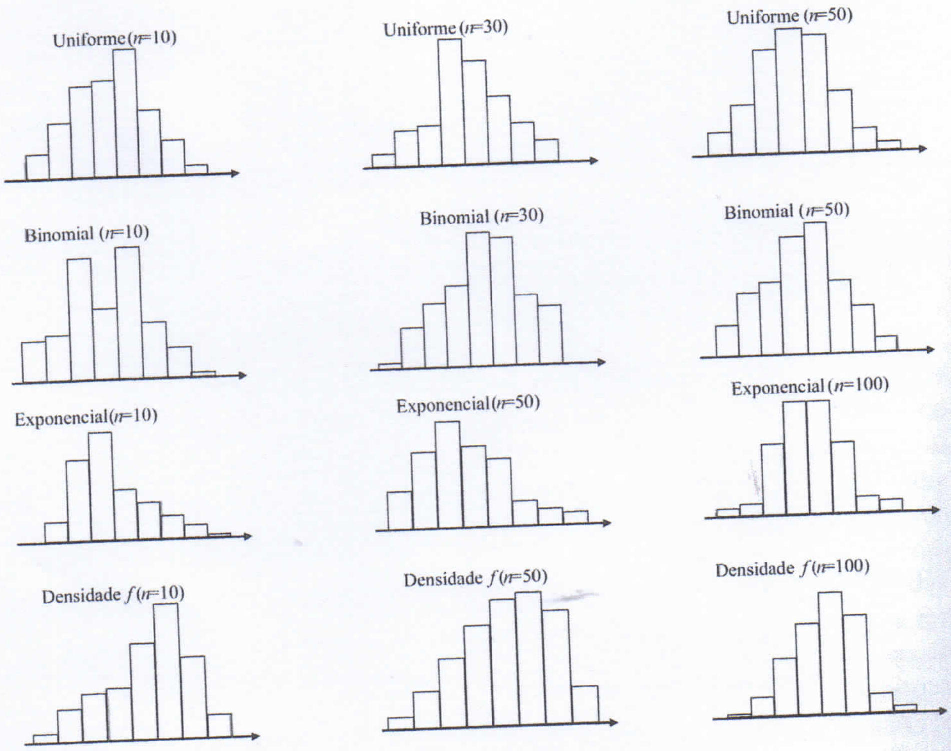


Figura 7.2: Efeito do tamanho da amostra sobre a distribuição de $\overline{X}$.

Histogramas de 100 valores de $\bar{X}$ para amostras de tamanho n das distribuições

Uniforme Discreta $P(X=i) = \frac{1}{10}$ $i = 1, 2, \ldots 10$

Binomial

Exponencial ($\alpha = 2$)

$$f(x) = \begin{cases} 1/8 & 0 < x < 4 \\ 1/2 & 4 \leq x < 5 \end{cases}$$

A velocidade de convergência é maior para distribuições simétricas.

Aplicações do TLC

1) Em uma cidade, a duração em minutos de conversas telefônicas em telefones públicos tem distribuição exponencial com parâmetro $\alpha = \frac{1}{3}$. Observando-se uma amostra aleatória de 50 dessas chamadas, qual é a probabilidade delas, em média, não ultrapassarem 4 minutos?

X - duração da conversa $\qquad X \sim Exp\left(\frac{1}{3}\right)$

$$f(x) = \frac{1}{3} e^{-\frac{1}{3}x}, \quad x \geq 0 \qquad \Rightarrow E(X) = \frac{1}{\alpha} = 3 = \mu$$

$$Var(X) = \frac{1}{\alpha^2} = 9 = \sigma^2$$

$X_1, X_2 \ldots X_{50}$ amostra aleatória da v.a. X

$$\bar{X} = \frac{\sum_{i=1}^{50} X_i}{50} \quad \text{média amostral}$$

Pelo teorema do limite central $\bar{X} \approx N\left(3, \frac{9}{50}\right)$

$$\Downarrow$$

$$Z = \frac{\bar{X} - 3}{\sqrt{9/50}} \approx N(0,1)$$

$$P(\bar{X} \leq 4) = P\left(\frac{\bar{X} - 3}{\sqrt{9/50}} \leq \frac{4-3}{\sqrt{9/50}}\right) \simeq P(Z \leq 2{,}38) = 0{,}991$$

2) Dimensionamento da amostra

Bussab e Morettin pag 287

X - variável aleatória $E(X) = \mu$ $Var(X) = \sigma^2$

Problema: estimar μ - média populacional através da média amostral $\bar{X}$ com base em uma amostra aleatória da v.a. X.

Deseja-se determinar o tamanho de amostra n de modo que a probabilidade de $\bar{X}$ não se distanciar de μ por mais que um erro ε seja γ $(0 < \gamma < 1)$.

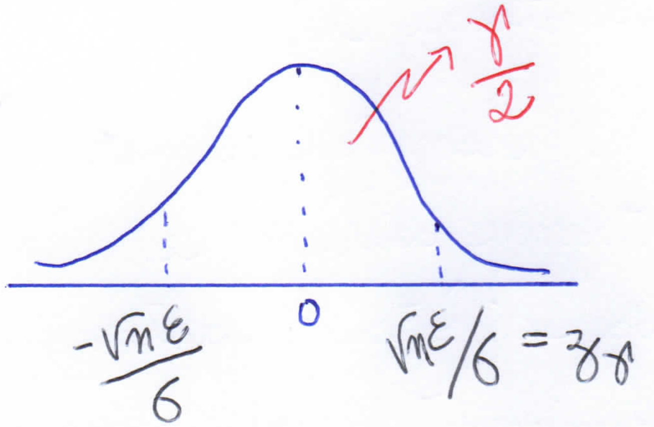
$P(|\bar{X} - \mu| \leq \varepsilon) = \gamma$ γ, ε dados (fixados)

Pelo TLC, $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$ para n grande.

$$P(-\varepsilon \leq \bar{X} - \mu \leq \varepsilon) = P\left(\frac{-\varepsilon}{\sigma/\sqrt{n}} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{\varepsilon}{\sigma/\sqrt{n}}\right)$$

$$\simeq P\left(-\frac{\sqrt{n}\,\varepsilon}{\sigma} \leq Z \leq \frac{\sqrt{n}\,\varepsilon}{\sigma}\right) = \gamma \qquad Z \sim N(0,1).$$

$$P\left(\overset{-z_\gamma}{-\frac{\sqrt{n}\,\varepsilon}{\sigma}} \le Z \le \overset{z_\gamma}{\frac{\sqrt{n}\,\varepsilon}{\sigma}}\right) = \gamma \qquad Z \sim N(0,1)$$



Dado γ, obtem-se z_γ da tabela da $N(0,1)$

$$\Rightarrow \frac{\sqrt{n}\,\varepsilon}{\sigma} = z_\gamma \Rightarrow n = \frac{\sigma^2 z_\gamma^2}{\varepsilon^2}$$

σ^2 é desconhecido.

Utilizar informações a priori sobre σ^2 ou estimar σ^2 através de uma amostra piloto.

3) Deseja-se estimar a nota média de Matemática dos alunos de uma grande escola. Por experiências anteriores sabe-se que o desvio padrão dessas notas é igual a 2 mas acredita-se que a média pode ter se modificado. Determine o tamanho de amostra necessário para que a estimativa da nota média não se desvie da verdadeira média por mais de 0,5 unidade com probabilidade 0,90.

X - nota de Matemática

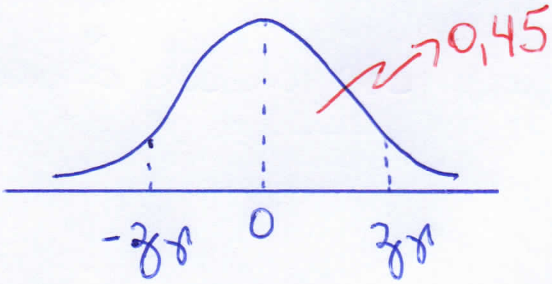
$\mu = E(X)$ - nota média desconhecida

$\bar{X}$ estimador de μ $\quad n = ?$

$P(|\bar{X} - \mu| \leq 1) = 0,9$ $\quad \sigma = 2$ $\quad \gamma = 0,9$

$\varepsilon = 0,5$

$$n = \frac{\sigma^2 z_\gamma^2}{\varepsilon^2}$$



$z_\gamma = 1,64$

$$n = \frac{4 \cdot 1,64^2}{0,5^2} = 43,03 \simeq 44$$