Matemática Computacional

Teórica 3

Departamento de Matemática
Instituto Superior de Engenharia do Porto

2º Semestre 21-22

Conteúdo

1 Teorema do limite central

2 Amostragem

3 Distribuições de Amostragem

Teorema do Limite Central

Teorema (TLC)

Se $X_1, X_2, \dots X_n$ *são* n *variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com média* μ *e variância finita* σ^2 *então a v.a.*

$$Y = \sum_{i=1}^{n} X_i \xrightarrow[n \to \infty]{D.} N\left(n\mu, n\sigma^2\right)$$

Nota: Na prática, considera-se uma boa aproximação quando $n \geq 30$.

Exemplo

Admite-se que o erro cometido (em mm) em cada operação de medição é uma v.a. com média $\mu = 0$ e desvio padrão $\sigma = 5$. Para a realização de um trabalho de medição realizaram-se 50 operações. Calcule a probabilidade do erro de medição acumulado nas 50 operações exceder 2 cm.

X_i - v.a. erro cometido na operação i, $i = 1, 2, \ldots, 50$

$Y = \sum_{i=1}^{50} X_i$ - v.a. erro cometido no total das 50 operações

Pelo teorema do limite central e sendo $n \geq 30$, tem-se

$$Y = \sum_{i=1}^{50} X_i \sim N\left(50 \times 0, 50 \times 5^2\right)$$

$P(Y > 20) = 1 - P(Y \leq 20) = 1 - \Phi\left(\frac{20-0}{35.35}\right) = 1 - \Phi\left(0.57\right) = 0.2843$

Aproximação da Binomial à Normal

Teorema (Moivre)

Se X é uma v.a. com distribuição Binomial, $X \sim B_i(n,p)$ de valor médio $\mu = np$ e variância $\sigma^2 = npq$ então

$$\frac{X - np}{\sqrt{npq}} \xrightarrow[n\to\infty]{D.} N(0,1)$$

Recorde-se que a distribuição Binomial pode ser aproximada pela distribuição de Poisson quando n é grande e $p \approx 0$.
Para valores de $p \approx 1/2$, o TLC garante uma boa aproximação da distribuição Binomial pela distribuição Normal.
Regra prática: quando $np > 5$ e $nq > 5$, podemos considerar boa a aproximação da distribuição Binomial pela distribuição Normal

Aproximação da Poisson à Normal

Teorema

Se X *é uma v.a. com distribuição de Poisson,* $X \sim P_o(\lambda)$ *então*

$$\frac{X-\lambda}{\sqrt{\lambda}} \xrightarrow[\lambda\to\infty]{D.} N(0,1)$$

Os teoremas anteriores permitem-nos aproximar uma variável discreta por uma contínua.
Neste caso, é necessário fazer-se uma *correção de continuidade*, que consiste em considerar

$$P(X=k) \approx P(k-0.5 < X < k+0.5)$$

$$P(X \leq k) \approx P(X < k+0.5)$$

Exemplo

Seja $X \sim B_i(15, 0.4)$. Calcule os valores exatos e os valores aproximados das seguintes probabilidades:

1. $P\left(X=5\right)$
2. $P\left(10 < X \leq 14\right)$

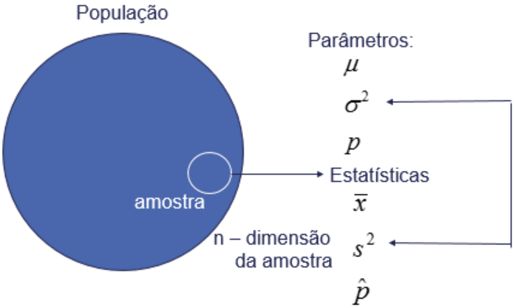
Resolução:

Os valores aproximados são calculados a partir da distribuição normal, $\frac{X-6}{1.9} \sim N(0,1)$

1. Valor "exato" de $P\left(X=5\right) = 0.1859$ (tabela da Binomial)
   Valor aproximado de $P\left(X=5\right) \approx P\left(4.5 < X < 5.5\right) = \Phi\left(\frac{5.5-6}{1.9}\right) - \Phi\left(\frac{4.5-6}{1.9}\right) = \Phi(-0.26) - \Phi(-0.79) = 0.1827$
2. Valor "exato" de $P\left(10 < X \leq 14\right) = 0.0093$
   Valor aproximado de $P\left(10 < X \leq 14\right) \approx P\left(10.5 < X < 14.5\right) = \Phi\left(\frac{14.5-6}{1.9}\right) - \Phi\left(\frac{10.5-6}{1.9}\right) = 0.0089$

Amostragem

Muitas aplicações da estatística a problemas reais consistem na recolha de amostras de populações e subsequente cálculo de certas medidas descritivas (média, proporção, etc) para a obtenção de informações/conclusões sobre as características das populações em estudo.



Amostra aleatória

Diz-se que $(X_1, X_2, \ldots, X_n)$ é uma amostra aleatória de dimensão n se as variáveis $X_i\,(i = 1, \ldots, n)$ são independentes e semelhantes, i.é., têm todas a mesma distribuição (igual à da população).

Nota: Antes da amostragem, as quantidades observáveis são variáveis aleatórias.
Depois da amostragem obtemos um conjunto de dados que passamos a representar por $x_1, x_2, \ldots, x_n$.
As amostras classificam-se em **grandes amostras** se $n \geq 30$ e em **pequenas amostras** se $n < 30$.

Estatística

Estatística é toda a função T da amostra aleatória, que não contenha parâmetros desconhecidos.

$T = T(X_1, X_2, \ldots, X_n)$ é uma variável aleatória função dos valores aleatórios.

A cada amostra observada $(x_1, x_2, \ldots, x_n)$ corresponde um valor numérico $t(x_1, x_2, \ldots, x_n)$.

Exemplos

- Média Amostral

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

- Variância Amostral

$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n-1}$$

Distribuições de Amostragem

Distribuição de amostragem

As estatísticas são, por definição, variáveis aleatórias que têm distribuições usualmente chamadas **distribuições de amostragem**.

Se $(X_1, X_2, \ldots, X_n)$ é uma amostra aleatória de uma população com função densidade de probabilidade $f(x|\theta)$ (θ - parâmetro desconhecido) então a distribuição de amostragem da estatística $T = T\left(X_1, X_2, \ldots, X_n\right)$ define-se a partir da distribuição conjunta $\prod_{i=1}^{n} f\left(x_i|\theta\right)$.

Distribuição da Média Amostral

População com distribuição Normal

Se $(X_1, X_2, \ldots, X_n)$ é uma amostra aleatória de uma população com distribuição Normal

$$X_i \sim N(\mu, \sigma^2)$$

então, pela aditividade da distribuição Normal, tem-se

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Distribuição da Média Amostral

Observe-se que:

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

$$V(\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Distribuição da Média Amostral

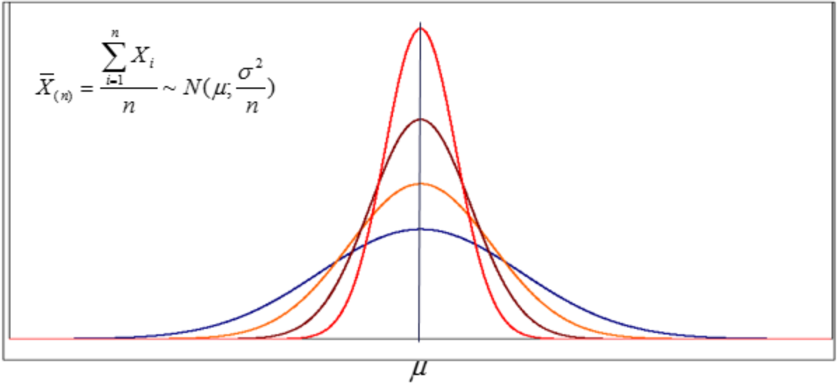
População com distribuição **NÃO** Normal

Se $(X_1, X_2, \ldots, X_n)$ é uma amostra aleatória de uma população que não tem distribuição Normal, com

$$E\left(X_i\right) = \mu \text{ e } V\left(X_i\right) = \sigma^2$$

então

- se $n \geq 30$, $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.
  Se σ^2 desconhecida, $\sigma^2 \approx s^2$
- se $n < 30$, $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ não tem distribuição normal



Obs: Quando a dimensão da amostra aumenta a dispersão diminui

Exemplo

Admite-se que a resistência à tração das peças produzidas por um fornecedor é uma v.a. $N(120kg, 25kg^2)$. Um cliente, interessado em realizar um grande negócio, combinou com o fornecedor a realização de um ensaio de tração a 40 peças escolhidas aleatoriamente.

Calcule a probabilidade de se realizar negócio sabendo que o comprador aceita o negócio caso se obtenha uma resistência média superior a 118 kg.

Exemplo: resolução

$\overline{X}$ - v. a. resistência média de 40 peças (Kg)
X_i - v. a. resistência da i-ésima peça (kg)

$X_i \sim N(120, 25),\ i = 1, 2, \ldots, 40$

$\overline{X} = \sum\limits_{i=1}^{40} \frac{X_i}{40} \sim N(120, \frac{25}{40})$

$$\begin{aligned} P(\overline{X} > 118) &= 1 - P(\overline{X} \leq 118) = 1 - \Phi\left(\frac{118 - 120}{\sqrt{25/40}}\right) \\ &= 1 - \Phi\left(-2.53\right) = 1 - \left(1 - \Phi\left(2.53\right)\right) \\ &= \Phi\left(2.53\right) = 0.9943 \end{aligned}$$

Distribuição da diferença entre duas médias amostrais

Duas populações com distribuições Normais

Se $(X_{A_1}, X_{A_2}, \ldots, X_{A_n})$ e $(X_{B_1}, X_{B_2}, \ldots, X_{B_m})$ são duas amostras aleatórias independentes de duas populações A e B com distribuição Normal, onde

$X_{A_i} \sim N(\mu_A, \sigma_A^2)$, $\overline{X}_A = \dfrac{\sum_{i=1}^{n} X_{A_i}}{n} \sim N\left(\mu_A, \dfrac{\sigma_A^2}{n}\right)$ e

$X_{B_i} \sim N(\mu_B, \sigma_B^2)$, $\overline{X}_B = \dfrac{\sum_{i=1}^{m} X_{B_i}}{m} \sim N\left(\mu_B, \dfrac{\sigma_B^2}{m}\right)$ então

$$\overline{X}_A - \overline{X}_B \sim N\left(\mu_A - \mu_B, \frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}\right)$$

Distribuição da diferença entre duas médias amostrais

Duas populações com distribuições **NÃO** Normais

Se $(X_{A_1}, X_{A_2}, \ldots, X_{A_n})$ e $(X_{B_1}, X_{B_2}, \ldots, X_{B_m})$ são duas amostras aleatórias independentes de duas populações A e B que não têm distribuição Normal, onde $E(X_{A_i}) = \mu_A$, $V(X_{A_i}) = \sigma_A^2$ e $E(X_{B_i}) = \mu_B$, $V(X_{B_i}) = \sigma_B^2$ então

- se $n \geq 30$ e $m \geq 30$, pelo teorema do limite central e aditividade da Normal, tem-se
  $\overline{X}_A - \overline{X}_B \sim N\left(\mu_A - \mu_B, \frac{\sigma_A^2}{n} + \frac{\sigma_B^2}{m}\right)$
- se $n < 30$ e $m < 30$, $\overline{X}_A - \overline{X}_B$ não tem distribuição Normal

Exemplo

Suponha que se realizou uma amostra de 40 peças a cada um de dois fornecedores, A e B. A produz peças cuja resistência é $N(120kg, 25kg^2)$ e B produz peças cuja resistência é uma v.a. com média $122kg$ e uma variância de $49kg^2$.
Qual a probabilidade da resistência média de 40 peças do fornecedor A exceder a resistência média de 40 peças do fornecedor B?

Resolução:

$$\overline{X}_A \sim N\left(120, \tfrac{25}{40}\right) \quad ; \quad \overline{X}_B \sim N\left(122, \tfrac{49}{40}\right)$$

$$Y = \overline{X}_A - \overline{X}_B \sim N\left(120 - 122, \tfrac{25}{40} + \tfrac{49}{40}\right) \Leftrightarrow Y \sim N(-2, 1.85)$$

$$P\left(\overline{X}_A > \overline{X}_B\right) = P\left(\overline{X}_A - \overline{X}_B > 0\right) = P(Y > 0) = 1 - P(Y \leq 0)$$
$$= 1 - \Phi\left(\frac{0 - (-2)}{\sqrt{1.85}}\right) = 1 - \Phi(1.36) = 0.0869$$

Distribuição da proporção amostral

De uma população, com uma proporção p de elementos que têm determinada característica em estudo, são recolhidas amostras aleatórias de dimensão n e calculada a correspondente proporção observada $\hat{p}_0$.
Seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória de dimensão $n > 30$ extraída, com reposição, de uma população.
Cada uma das variáveis X_i tem uma distribuição de Bernoulli de parâmetro p.

$$X_i \sim Be(p)$$
$$E\left(X_i\right) = p$$
$$V\left(X_i\right) = p(1-p)$$

Distribuição da proporção amostral

Seja $X = \sum_{i=1}^{n} X_i$ a v.a. que representa o nº de elementos da população que têm a característica em estudo.
Sabemos que $X \sim B_i(n, p)$.
Como $n > 30$ podemos assegurar, pelo teorema do limite central, que X tem uma distribuição aproximadamente Normal

$$X = \sum_{i=1}^{n} X_i \sim N\left(np, npq\right)$$

Proporção amostral

Definimos $\hat{P}$ a v.a. que representa a proporção (amostral) de elementos possuidores da característica em estudo, em n elementos amostrados

$$\hat{P} = \frac{X}{n} \sim N\left(p, \frac{pq}{n}\right)$$

Observe-se que

$$E(\hat{P}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{np}{n} = p$$

$$V(\hat{P}) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}V(X) = \frac{npq}{n^2} = \frac{pq}{n}$$

Exemplo

Um fornecedor produz componentes com uma taxa de 5% de defeituosos. Um comprador retira uma amostra de 50 peças do total produzido e calcula a percentagem de componentes defeituosos. Calcule a probabilidade de na amostra, se observar uma percentagem de componentes defeituosos superior a 6%.

Resolução:

$\hat{P}$ - v.a que representa a proporção de peças defeituosas observada numa amostra de 50 peças

$p = 0.05$ - proporção de peças defeituosas produzidas pelo fornecedor

$\hat{P} \sim N\left(0.05, \frac{0.05 \times 0.95}{50}\right) \Leftrightarrow \hat{P} \sim N\left(0.05, 0.03^2\right)$

$P(\hat{P} > 0.06) = 1 - P(\hat{P} \leq 0.06) = 1 - P\left(Z \leq \frac{0.06 - 0.05}{0.03}\right) = 1 - \Phi(0.33) = 0.3707$

Distribuição da diferença entre duas proporções amostrais

Distribuição da diferença entre duas proporções

Se $(X_{A_1}, X_{A_2}, \ldots, X_{A_n})$ e $(X_{B_1}, X_{B_2}, \ldots, X_{B_m})$ são duas amostras aleatórias independentes de duas populações A e B, p_A e p_B as respetivas proporções de elementos possuidores de uma característica em estudo e $\hat{P}_A \sim N\left(p_A, \frac{p_A q_A}{n}\right)$ e $\hat{P}_B \sim N\left(p_B, \frac{p_B q_B}{m}\right)$ as v.a. proporções amostrais então

$$\hat{P}_A - \hat{P}_B \sim N\left(p_A - p_B, \frac{p_A q_A}{n} + \frac{p_B q_B}{m}\right)$$

Exemplo

Na tabela seguinte indica-se a percentagem de peças defeituosas produzidas por 2 fornecedores

Fornecedor	Percentagem de defeituosas
A	6%
B	4%

Qual a probabilidade da percentagem de peças defeituosas produzidas pelo fornecedor B ser superior à do fornecedor A, quando são retiradas duas amostras aleatórias, de 60 peças de cada fornecedor?

Exemplo: resolução

$\hat{P}_A$ - proporção de peças defeituosas produzidas pelo fornecedor A, numa amostra aleatória de 60 peças

$$\hat{P}_A \sim N\left(0.06, \frac{0.06 \times 0.94}{60}\right)$$

$\hat{P}_B$ -proporção de peças defeituosas produzidas pelo fornecedor B, numa amostra aleatória de 60 peças.

$$\hat{P}_B \sim N\left(0.04, \frac{0.04 \times 0.96}{60}\right)$$

$$D = \hat{P}_B - \hat{P}_A \sim N\left(0.04 - 0.06, \tfrac{0.04 \times 0.96}{60} + \tfrac{0.06 \times 0.94}{60}\right)$$
$$D \sim N\left(-0.02, 0.04^2\right)$$

$$P\left(D > 0\right) = 1 - P(D \leq 0) = 1 - P\left(Z \leq \frac{0 - (-0.02)}{0.04}\right) = 1 - \Phi(0.5) = 0.3085$$