

Licenciatura em Engenharia Informática  
(LEI/LEICE/LEIRP)  
Ano Letivo 2022/23

---

# MÉTODOS ESTATÍSTICOS

RESUMO DE APOIO ÀS AULAS

DEOLINDA M. L. D. RASTEIRO



INSTITUTO SUPERIOR DE ENGENHARIA DE COIMBRA  
DEPARTAMENTO DE FÍSICA E MATEMÁTICA

1. Probabilidades	3
2. Variáveis Aleatórias e Distribuições de Probabilidade Discretas	9
3. Variáveis Aleatórias e Distribuições de Probabilidade Contínuas	26
4. Amostragem e Distribuições Amostrais	35
5. Estimação	38
6. Teste de Hipóteses Paramétricos	43

# 1 Probabilidades

## 1.1 Experiência aleatória, espaço de resultados, acontecimentos

### Definição

*Experiência aleatória* é um qualquer processo ou conjunto de circunstâncias capaz de produzir pelo menos dois resultados, com incerteza quanto ao que ocorrerá.

#### Características principais:

- Possibilidade de repetição;
- Carácter imprevisível;
- Apresentam regularidade estatística.

### Definição

*Espaço de resultados* ou *espaço fundamental* é o conjunto de todos os resultados possíveis da experiência aleatória. Denota-se por  $\Omega$ .

### Definições

Um *acontecimento* (ou *evento*) é um subconjunto de  $\Omega$ .

*Acontecimento elementar* é um subconjunto singular de  $\Omega$ .

- $\Omega$  é denominado *acontecimento certo* (realiza-se sempre);
- $\emptyset$  é denominado *acontecimento impossível*;
- $\overline{A}$  é denominado *acontecimento complementar* de  $A$ .

### Operações e relações entre acontecimentos

1.  $A \subset B$ : a realização de  $A$  implica a realização de  $B$ ;
2.  $A = B$ :  $A \subset B$  e  $B \subset A$ ; ( $A$  e  $B$  dizem-se idênticos )
3.  $A \cap B$  (Acontecimento Intersecção):  $A$  e  $B$  realizam-se conjuntamente;
  - Se  $A \cap B = \emptyset$  então  $A$  e  $B$  dizem-se **mutuamente exclusivos, disjuntos** ou **incompatíveis**.
4.  $A \cup B$  (Acontecimento Reunião):  $A$  ou  $B$  se realizam (o resultado da experiência aleatória pertence a pelo menos um dos conjuntos);
5.  $A \setminus B$  (Acontecimento Diferença):  $A$  realiza-se e  $B$  não se realiza;
  - $\overline{A} = \Omega \setminus A$
  - $A \setminus B = A \cap \overline{B}$

## Propriedades

1.  $A \cap \overline{A} = \emptyset$

2.  $A \cup \overline{A} = \Omega$

3. **Comutativa**

$$A \cap B = B \cap A$$

$$A \cup B = B \cup A$$

4. **Associativa**

$$A \cap (B \cap C) = (A \cap B) \cap C$$

$$A \cup (B \cup C) = (A \cup B) \cup C$$

5. **Distributiva**

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

6.  $A \cap \Omega = A$ ;  $A \cup \emptyset = A$

7.  $A \cup \Omega = \Omega$ ;  $A \cap \emptyset = \emptyset$

8.  $A \subset B \Rightarrow A \cup B = B$  e  $A \cap B = A$

9. **Leis de De Morgan**

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

## 1.2 Definição de probabilidade

### Definição [Clássica]

Admita-se que  $\Omega$  é um espaço finito e que todos os acontecimentos elementares são equipossíveis. A probabilidade de um acontecimento (qualquer subconjunto de  $\Omega$ )  $A$  se realizar é dada por

$$P(A) = \frac{\#A}{\#\Omega}.$$

- $P(\Omega) = 1$
- $P(\emptyset) = 0$
- $\forall A \in \Omega, 0 \leq P(A) \leq 1.$

### Definição [Axiomática]

Seja  $\Omega$  um espaço de resultados e  $A$  e  $A_i, i = 1, 2, \dots$ , acontecimentos quaisquer de  $\Omega$ .

Uma probabilidade é uma aplicação  $P$  que satisfaz os seguintes axiomas:

(1)  $0 \leq P(A) \leq 1$

(2)  $P(\Omega) = 1$

(3) Para o acontecimento reunião  $\bigcup_{i=1}^{+\infty} A_i$  de  $\Omega$ ,

$$P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} P(A_i), \text{ se } A_i \cap A_j = \emptyset \text{ para } i \neq j.$$

### Nota

A partir de  $\Omega$  é possível formar várias famílias de subconjuntos deste espaço (**pensemos no caso de  $\Omega$  não ser finito**). A definição axiomática de probabilidade, definida anteriormente no domínio  $\Omega$ , estende-se à família (chamemos-lhe  $F$ ) de todos os subconjuntos de  $\Omega$ , que verifica:

(1)  $\Omega \in F$ ;

(2) Se  $A \in F$  então  $\overline{A} \in F$ ;

(3) Se  $A_i \in F, i \in \mathbb{N}$ , então  $\bigcup_{i=1}^{+\infty} A_i \in F$ .

## Propriedades de uma probabilidade

1.  $P(\emptyset) = 0$

2. Se  $A_i \in \Omega$ ,  $i = 1, \dots, n$ , e  $A_i \cap A_j = \emptyset$  para  $i \neq j$ , então

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

**Caso particular:**

$$\text{Se } A \text{ e } B \in \Omega \text{ e } A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B).$$

3. Se  $A \text{ e } B \in \Omega$ ,  $P(A \setminus B) = P(A) - P(A \cap B)$

4. Se  $A \text{ e } B \in \Omega$  e  $B \subset A \Rightarrow P(B) \leq P(A)$

5.  $A \in \Omega$ ,  $P(\overline{A}) = 1 - P(A)$

6.  $A, B \in \Omega$ ,  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

7. Se  $A_i \in \Omega$ ,  $i = 1, \dots, n$ ,

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i);$$

8. Se  $A_i \in \Omega$ ,  $i \in \mathbb{N}$ , e  $A_i$  é uma sucessão monótona,

$$P\left(\lim_{i \rightarrow +\infty} A_i\right) = \lim_{i \rightarrow +\infty} P(A_i).$$

### 1.3 Probabilidade condicionada

**Definição** Sejam  $A$  e  $B$  acontecimentos de  $\Omega$  com  $P(B) > 0$ . A probabilidade de  $A$  condicionada por  $B$ ,  $P(A/B)$ , é dada por

$$P(A/B) = \frac{P(A \cap B)}{P(B)}.$$

**Teorema [probabilidade composta]** Se  $A$  e  $B$  são acontecimentos de  $\Omega$  tais que  $P(A)P(B) > 0$ , então

$$P(A \cap B) = P(A/B)P(B) = P(B/A)P(A).$$

**Generalização:** Sejam  $A_1, A_2, \dots, A_n$  acontecimentos de um mesmo espaço  $\Omega$ , tais que  $P(A_1 \cap A_2 \cap \dots \cap A_n) > 0$ . Então

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \dots P(A_n/A_1 \cap A_2 \cap \dots \cap A_{n-1})$$

## 1.4 Acontecimentos independentes

**Definição [Independência]** Os acontecimentos  $A$  e  $B$  dizem-se *independentes* se e só se

$$P(A \cap B) = P(A)P(B).$$

**Consequência:** Sejam  $A$  e  $B$  acontecimentos de  $\Omega$  tais que  $P(A)P(B) > 0$ .  $A$  e  $B$  são independentes se e só se

$$P(A/B) = P(A).$$

**Nota:** Não confundir acontecimentos **disjuntos** com acontecimentos **independentes**.

## 1.5 Probabilidade total. Teorema de Bayes

**Teorema [probabilidade total]** Sejam  $A_1, A_2, \dots, A_n$  acontecimentos de  $\Omega$  tais que

$$\begin{aligned} A_i \cap A_j &= \emptyset, \quad i \neq j && (\text{disjuntos}) \\ \bigcup_{i=1}^n A_i &= \Omega && (\text{exaustivos}) \end{aligned}$$

Seja  $B$  um acontecimento qualquer. Tem-se

$$P(B) = \sum_{i=1}^n P(B/A_i) P(A_i).$$

**Teorema [Bayes]** Sejam  $A_1, A_2, \dots, A_n$  acontecimentos de  $\Omega$  tais que

$$\begin{aligned} A_i \cap A_j &= \emptyset, \quad i \neq j \\ \bigcup_{i=1}^n A_i &= \Omega. \end{aligned}$$

Seja  $B$  um acontecimento qualquer, com  $B \neq \emptyset$ . Tem-se

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{\sum_{i=1}^n P(B/A_i) P(A_i)}, \quad i = 1, \dots, n.$$

## Exercícios Complementares

1. Uma companhia de seguros classifica os seus segurados em três categorias: baixo risco, risco médio e risco elevado. Os seus registos indicam que a probabilidade de um segurado se envolver em pelo menos um acidente, por ano, é 0.01, 0.10, e 0.25 se o segurado pertence à categoria de baixo, médio ou risco elevado, respectivamente. Admita que a probabilidade de um segurado ser classificado na categoria de baixo risco é de 0.1 enquanto que na de risco médio é 0.6.
  - (a) Qual a probabilidade de, num ano, um dos segurados tenha pelo menos um acidente?
  - (b) Sabendo que um dos segurados teve pelo menos um acidente no último ano, qual a probabilidade de pertencer à categoria de risco elevado?
  - (c) Sabendo que um dos segurados não teve acidentes no último ano, qual a probabilidade dele pertencer à categoria de risco médio?
2. Numa fábrica as máquinas I, II e III produzem peças do mesmo comprimento na proporção de 35:25:40. Sabe-se que 2% das peças produzidas pela máquina I são defeituosas e 1% das peças produzidas pela máquina II são defeituosas. Sabe-se ainda que 1.2% das peças são produzidas pela máquina III e são defeituosas.
  - (a) Se for escolhida aleatoriamente uma peça da produção da fábrica, qual a probabilidade de ser defeituosa?
  - (b) Se for seleccionada uma peça defeituosa, qual a probabilidade de ter sido produzida pela máquina II?



## 2 Variáveis Aleatórias e Distribuições de Probabilidade Discretas

### 2.1 Introdução

**Definição** Seja  $\Omega$  um espaço de resultados associado a uma experiência aleatória. Designa-se por *variável aleatória* (v.a.) uma função (correspondência unívoca)  $X$  de domínio  $\Omega$  e contradomínio em  $\mathbb{R}$ .

$$\begin{aligned} X : \Omega &\longrightarrow \mathbb{R} \\ w &\longrightarrow X(w) \end{aligned}$$

#### Exemplos de Ilustração

1. Lançamento de duas moedas equilibradas, uma a seguir à outra, anotando as faces que ficam voltadas para cima.

$$\Omega = \{(C, C), (C, \overline{C}), (\overline{C}, C), (\overline{C}, \overline{C})\}, \text{ com } C : \text{“saída de cara”}.$$

Seja  $X$  a v.a. que representa (por exemplo) o **número** de caras obtidas no lançamento das duas moedas.

$$X(w) \leftarrow \text{número de caras no elemento } w \text{ de } \Omega.$$

2. Lançamento de dois dados, anotando os números de pontos das faces que ficam voltadas para cima.

$$\Omega = \{(i, j), i, j = 1, 2, 3, 4, 5, 6\}.$$

Seja  $X$  a v.a. que representa a **soma** dos pontos obtidos.

3. Medição da altura de uma pessoa, escolhida ao acaso.

$\Omega$  é o conjunto de todas as alturas atribuíveis a uma pessoa.

Seja  $X$  a v.a. que representa a **altura** de uma pessoa.

**Observação** Em algumas situações o conjunto de valores que uma v.a. toma (contradomínio) confunde-se com o próprio  $\Omega$ ; noutras, os valores da v.a. distinguem-se claramente dos elementos de  $\Omega$ . Embora da experimentação nem sempre resultem **valores numéricos**, o conceito de variável aleatória permite associar valores numéricos (números ou intervalos reais) aos elementos de  $\Omega$ , relação esta com grande interesse prático.

Para  $w \in \Omega$ , como calcular a probabilidade de ocorrência de  $X(w)$ ? (“deixámos”  $\Omega$ , estamos em  $\mathbb{R}$  !!!)

Uma variável aleatória  $X : \Omega \longrightarrow \mathbb{R}$  é tal que:

- $\forall A \subseteq \Omega, X(A) = \{X(w) : w \in A\}$  (imagem de A por X)
  - $\forall E \subseteq \mathbb{R}, X^{-1}(E) = \{w \in \Omega : X(w) \in E\}$  (imagem inversa de E por X)
- Note que se  $w \in X^{-1}(E) \Rightarrow X(w) \in E$ .

(Uma v.a. é uma aplicação que estabelece uma relação entre subconjuntos de  $\Omega$  e subconjuntos de  $\mathbb{R}$  e **vice-versa**)

### Propriedade

Sendo  $E$  um qualquer subconjunto de  $\mathbb{R}$  e  $X$  uma variável aleatória,  $X^{-1}(E)$  é um subconjunto de  $\Omega$ .

Como  $P$  está definida neste domínio, podemos sempre calcular  $P(X^{-1}(E))$ ! É assim introduzida em  $\mathbb{R}$  uma **probabilidade**, denotada por  $P_X$ , definida por

$$P_X(B) = P(X^{-1}(B)), \forall B \subseteq \mathbb{R}.$$

A  $P_X$  é usual chamar-se lei de probabilidade da v.a.  $X$ .

### Notação

$$\begin{aligned} X^{-1}(B) &= \{w \in \Omega : X(w) \in B\} \equiv \{X \in B\} \\ \Rightarrow P_X(B) &= P(X^{-1}(B)) = P(X \in B) \end{aligned}$$

Na nossa incursão pelas variáveis aleatórias vamos considerar dois tipos de variáveis: as **discretas** e as (absolutamente) **contínuas** (Capítulo 3). Esta classificação é ditada pelos próprios fenómenos aleatórios em estudo e conseqüente leitura, mas também porque estas variáveis obedecem a certas regras (matemáticas) para que possam ser classificadas deste modo.

## 2.2 Variáveis Aleatórias Discretas

**Definição** Uma v.a.  $X : \Omega \longrightarrow \mathbb{R}$  diz-se discreta se  $\exists S \subset \mathbb{R} : S$  é finito ou infinito numerável tal que  $P_X(S) = P(X \in S) = 1$ .

Ao menor subconjunto de  $\mathbb{R}$  que tem probabilidade 1 chamamos *Suporte* de  $X$ .

**Definição** Seja  $X$  uma v.a. discreta. Chama-se **função de probabilidade de  $X$**  à função  $p : \mathbb{R} \longrightarrow [0, 1]$  definida por

$$p(x) = P(X = x), \forall x \in \mathbb{R}.$$

### Propriedades

$$(1) \quad 0 \leq p(x) \leq 1, \forall x \in \mathbb{R}$$

$$(2) \quad \sum_x p(x) = 1.$$

**Observação** Se  $X$  tem suporte  $S$  então  $p(x) = \begin{cases} P(X = x) & \text{se } x \in S \\ 0 & \text{se } x \in \overline{S} \end{cases}$ .

### Notas

- O conhecimento da função de probabilidade de  $X$ ,  $p$ , implica o conhecimento da lei de probabilidade de  $X$ ,  $P_X$ , e por isso  $p$  também é designada por lei de probabilidade de  $X$ .
- Identificamos sempre as v.a.'s por letras maiúsculas ( $X$ ) e as suas concretizações por letras minúsculas ( $x$ ).

**Definição** Seja  $X$  uma v.a. discreta. Chama-se *função distribuição de  $X$*  à função  $F : \mathbb{R} \rightarrow [0, 1]$  definida por

$$F(x) = P(X \leq x), \quad \forall x \in \mathbb{R}.$$

Se  $X$  tem suporte  $S$  então

$$F(x) = \sum_{\mathbf{a} \in S \cap ]-\infty, x]} P(X = \mathbf{a}).$$

↓

“valores do suporte de  $X$  e que são inferiores ou iguais a  $x$ ”

### Propriedades da função distribuição de uma v.a. discreta

- (1)  $\forall x \in \mathbb{R}, 0 \leq F(x) \leq 1$  ( $F$  é uma função limitada)
- (2)  $\forall a, b \in \mathbb{R}, \text{ se } a < b \Rightarrow F(a) \leq F(b)$  ( $F$  é não decrescente)
- (3)  $\lim_{x \rightarrow -\infty} F(x) = 0$  e  $\lim_{x \rightarrow +\infty} F(x) = 1$
- (4)  $\forall a \in \mathbb{R}, \lim_{x \rightarrow a^+} F(x) = F(a)$  ( $F$  é contínua à direita)
- (5)  $F$  admite um número finito ou infinito numerável de pontos de des-continuidade (pontos do Suporte de  $X$ )
- (6) A cada v.a.  $X$  corresponde uma e uma só função distribuição
- (7) Para  $\forall a, b \in \mathbb{R}$ ,

$$\begin{aligned} P(a < X < b) &= P(X < b) - P(X \leq a) \\ &= \lim_{x \rightarrow b^-} F(x) - F(a); \\ P(a \leq X \leq b) &= F(b) - F(a) - P(X = a); \\ P(a \leq X < b) &= F(b) - P(X = b) - F(a) + P(X = a); \\ &\vdots \end{aligned}$$

### Independência de Variáveis Aleatórias

**Definição** Sejam  $X$  e  $Y$  duas variáveis aleatórias (discretas ou contínuas).  $X$  e  $Y$  dizem-se independentes se e só se

$$\forall x, y \in \mathbb{R}, P(X \leq x \cap Y \leq y) = P(X \leq x) \times P(Y \leq y).$$

### Generalização

As  $n$  ( $n \geq 2$ ) variáveis aleatórias  $X_1, X_2, \dots, X_n$  dizem-se independentes se e só se

$$\forall x_1, x_2, \dots, x_n \in \mathbb{R}, P(X_1 \leq x_1 \cap X_2 \leq x_2 \cap \dots \cap X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i).$$

### Parâmetros de localização e de dispersão

**Definição** Seja  $X$  uma v.a. discreta de suporte  $S$ . A **esperança matemática de  $X$**  (ou **valor médio de  $X$**  ou **valor esperado de  $X$** ), caso exista, é definida por

$$E(X) = \sum_{x \in S} xP(X = x).$$

### Notas

- A existência do valor médio de uma v.a.  $X$  depende da convergência da série anterior.
- Se a v.a.  $X$  é discreta com suporte finito, então  $E(X)$  existe sempre!
- O valor médio de uma v.a. é um parâmetro de localização (tendência central).

**Definição** Se  $E(X) = 0$  diz-se que a v.a.  $X$  é centrada.

### Propriedades da esperança matemática

1. Seja  $X = c$ ,  $c \in \mathbb{R}$ .  $E(X) = E(c) = c$ .
2. Seja  $X$  uma v.a. discreta de suporte  $S$  e  $g : \mathbb{R} \rightarrow \mathbb{R}$  é tal que  $g(X)$  é uma v.a. discreta. Tem-se

$$E[g(X)] = \sum_{x \in S} g(x)P(X = x).$$

- Seja  $Y = g(X) = aX + b$ ,  $\forall a, b \in \mathbb{R}$ . Se  $E(X)$  existe então  $E(Y) = aE(X) + b$ .

- Seja  $k \in \mathbb{N}$  (arbitrariamente fixo) e  $g(X) = X^k$ .

$$E(X^k) = \sum_{x \in S} x^k P(X = x) \quad (\text{momento de ordem } k \text{ de } X).$$

3. Se  $E(X)$  existe então  $|E(X)| \leq E(|X|)$ .
4. Sejam  $X_1, X_2, \dots, X_n$  ( $n \in \mathbb{N}$ ) v.a.'s todas definidas sobre o mesmo espaço  $\Omega$  e tais que  $E(X_i)$  existe para  $i = 1, 2, \dots, n$ . Então

$$E\left[\sum_{i=1}^n a_i X_i + b\right] = \sum_{i=1}^n a_i E(X_i) + b,$$

para  $\forall a_1, a_2, \dots, a_n, b \in \mathbb{R}$ . (**Linearidade da esperança**)

$$\bullet E \left[ \sum_{i=1}^n X_i \right] = \sum_{i=1}^n E(X_i).$$

**Definição** Seja  $X$  uma v.a. discreta de suporte  $S$  e esperança  $E(X) = \mu$ . A **variância de  $X$** , denotada por  $V(X)$ ,  $Var(X)$  ou  $\sigma^2$ , é definida por

$$\begin{aligned} V(X) &= E[(X - \mu)^2] \\ &= \sum_{x \in S} (x - \mu)^2 P(X = x) \end{aligned}$$

**Definição** Chama-se **desvio-padrão de  $X$**  à raiz quadrada positiva da variância de  $X$ , e denota-se por  $\sigma(X) = \sigma = \sqrt{V(X)}$ .

**Nota** A variância e o desvio padrão de uma v.a.  $X$  são ambos parâmetros de dispersão; no entanto, o desvio padrão é mais usado, uma vez que vem expresso nas mesmas unidades de  $X$ .

### Propriedades da variância

1. Seja  $X$  uma v.a. tal que  $E(X)$  e  $E(X^2)$  existem. Então

$$V(X) = E(X^2) - E^2(X).$$

2. Se  $X = c$ ,  $c \in \mathbb{R}$ , então  $V(X) = 0$ .

3. Seja  $X$  uma v.a. tal que  $E(X)$  e  $E(X^2)$  existem e  $a, b \in \mathbb{R}$ . Então

$$V(aX + b) = a^2 V(X).$$

- $V(X + b) = V(X)$ ;
- $V(aX) = a^2 V(X)$ ;
- $V(-X) = V(X)$ .

### Outros parâmetros: Mediana e Quantis

**Definição** Chama-se **mediana** de  $X$  ao número real  $M_d$  tal que

$$F(M_d) = F(M_d^+) \geq 0.5 \quad \wedge \quad F(M_d^-) \leq 0.5$$

**Nota** A mediana é um parâmetro de localização, alternativo (ou complementar) ao valor esperado.

**Definição** Seja  $p \in ]0, 1[$ . Chama-se **quantil de ordem  $p$  de  $X$**  ao número real  $x_p$  tal que

$$F(x_p^+) \geq p \quad \wedge \quad F(x_p^-) \leq p$$

## 2.3 Distribuições Especiais Discretas

Vimos que associado a uma variável aleatória temos um modelo matemático que descreve (e resume) o comportamento dessa mesma variável e, conseqüentemente, o fenómeno aleatório em estudo. Deste modelo fazem parte a sua Distribuição de probabilidade (caracterizada pelas funções de probabilidade ou distribuição), os seus parâmetros de localização e de dispersão.

Nesta secção vamos conhecer alguns modelos, Leis, Famílias de distribuições, que são especiais porque uma mesma lei pode ser usada numa grande variedade de aplicações que envolvem fenómenos aleatórios.

Notemos a utilidade que uma *expressão comum*, isto é, que um *modelo* para a distribuição de variáveis aleatórias associadas a diferentes experiências aleatórias tem, pois simplifica o seu estudo e o nosso trabalho!!!

Estas distribuições estão bem estudadas e tabeladas. De entre um vasto leque de distribuições especiais, estudaremos apenas algumas, discretas nesta secção e contínuas no Capítulo 3, essenciais num estudo introdutório a esta área.

### 2.3.1 Distribuição de Bernoulli

Seja  $A$  um acontecimento associado a uma determinada experiência aleatória, tal que  $P(A) = p$ , com  $p \in ]0, 1[$ .

Considere-se a variável aleatória  $X$  definida do modo seguinte

$$X = \begin{cases} 1 & \text{se } A \text{ ocorre} \\ 0 & \text{se } A \text{ não ocorre} \end{cases}.$$

A variável assim definida é uma v.a. discreta de Suporte  $S = \{0, 1\}$  e função de probabilidade

$$p(x) = \begin{cases} 1-p & \text{se } x=0 \\ p & \text{se } x=1 \\ 0 & \text{se } x \in \overline{S} \end{cases} = \begin{cases} p^x (1-p)^{1-x} & \text{se } x \in \{0, 1\} \\ 0 & \text{se } x \in \overline{S} \end{cases}.$$

- A função (lei) de probabilidade assim definida é denominada por **lei de Bernoulli de parâmetro  $p$** , denotada por  $\mathcal{B}(p)$  (ou **Bernoulli( $p$ )**).
- Diz-se que a v.a.  **$X$  segue uma lei de Bernoulli** (ou  **$X$  tem distribuição de Bernoulli**) **de parâmetro  $p$** , e escreve-se simbolicamente  $X \sim \mathcal{B}(p)$ .

**Nota** A distribuição de Bernoulli está associada a experiências aleatórias com apenas dois resultados possíveis, usualmente denominados por **sucesso** e **insucesso**, com probabilidade de ocorrer sucesso  $p$  (o seu parâmetro!) e insucesso  $q = 1 - p$ . Uma experiência aleatória assim caracterizada chama-se **prova (ou tentativa) de Bernoulli**.

**Propriedades** Se  $X \sim \mathcal{B}(p)$  então

$$\begin{aligned} E(X) &= p \\ V(X) &= p(1-p) \end{aligned}$$

### 2.3.2 Distribuição Binomial

Seja  $A$  um acontecimento associado a uma determinada experiência aleatória, tal que  $P(A) = p$ , com  $p \in ]0, 1[$ .

Suponha agora que esta mesma experiência é realizada  $n$  vezes, com  $n > 1$ , sempre nas mesmas condições (note que os resultados das sucessivas experiências são independentes e logo  $P(A) = p$  em qualquer realização).

Considere-se a variável aleatória

**X: número de vezes que ocorre o acontecimento A, nas n realizações da experiência aleatória**

#### Observações

- O suporte de  $X$  é  $S = \{0, 1, \dots, n\}$
- A probabilidade (por exemplo) de ocorrer exactamente  $x$  vezes o acontecimento  $A$ , e de seguida  $n - x$  vezes  $\bar{A}$ , é dada por

$$P(\underbrace{A \cap \dots \cap A}_{x \text{ vezes}} \cap \underbrace{\bar{A} \cap \dots \cap \bar{A}}_{n-x \text{ vezes}}) = \underbrace{P(A) \dots P(A)}_{x \text{ vezes}} \underbrace{P(\bar{A}) \dots P(\bar{A})}_{n-x \text{ vezes}} = p^x (1 - p)^{n-x}$$

- O número de maneiras distintas de ocorrer  $x$  vezes o acontecimento  $A$ , nas  $n$  realizações, é  $C_x^n = \binom{n}{x} = \frac{n!}{(n-x)!x!}$
- A probabilidade de ocorrer exactamente  $x$  vezes o acontecimento  $A$ , nas  $n$  realizações, é dada por

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x \in \{0, 1, \dots, n\}.$$

$X$  é uma v.a. discreta de **suporte**  $S = \{0, 1, \dots, n\}$  e **função de probabilidade** dada por

$$p(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x} & \text{se } x \in \{0, 1, \dots, n\} \\ 0 & \text{se } x \in \bar{S} \end{cases}.$$

- $P(X = x) > 0, \forall x \in \{0, 1, \dots, n\}$
- $\sum_{x=0}^n P(X = x) = 1$

Diz-se que a v.a. **X segue uma lei Binomial** (ou **X tem distribuição Binomial**) de parâmetros **n** e **p**, e escreve-se simbolicamente  $\mathbf{X} \sim \mathcal{B}(n, p)$ .

Diz-se que a v.a. **X segue uma lei Binomial** (ou **X tem distribuição Binomial**) de parâmetros **n** e **p**, e escreve-se simbolicamente  $\mathbf{X} \sim \mathcal{B}(n, p)$ . (Interpretação de  $Y$  ao cuidado do aluno!)

**Observação** A lei Binomial descreve a contagem do número de *sucessos* em  $n$  repetições, independentes, de uma prova de Bernoulli com probabilidade de sucesso  $p$ . Note que se  $X_i$  representar o número de sucessos (0 ou 1) obtidos na prova  $i$ , então  $X_i \sim \mathcal{B}(p)$ ,  $i = 1, 2, \dots, n$ , e o número de sucessos obtidos nas  $n$  provas é a v.a.

$$X = \sum_{i=1}^n X_i \sim \mathcal{B}(n, p).$$

Uma das aplicações mais interessantes da lei Binomial é a contagem do número de elementos de determinado tipo numa amostra de dimensão  $n$ , quando seleccionados com reposição.

**Nota** A família de distribuições Binomial depende apenas dos parâmetros  $n$  e  $p$ . Esta está implementada na maioria das máquinas de calcular, assim como tabelada nos livros da área. Em particular, nas aulas de ME podem (ou devem, em falta de alternativa) ser consultadas as tabelas da disciplina onde constam as probabilidades

$$F(x) = P(X \leq x) = \sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k} \text{ para } x \in \{0, 1, \dots, n\}.$$

### 2.3.3 Distribuição Hipergeométrica

Suponha que se tem um número finito de  $N$  objectos, dos quais  $M$  são de um tipo (*sucesso*) e os restantes  $N - M$  de outro tipo (*insucesso*).

Considere a experiência aleatória: extracção, ao acaso, de um objecto.

Suponha agora que esta mesma experiência é realizada  $n$  vezes, isto é, extracção sucessiva, e sem reposição, de  $n$  dos  $N$  objectos (note que os resultados das sucessivas experiências não são independentes).

Seja

**X: número de objectos do tipo sucesso obtidos nas  $n$  extracções**

**Observações**

- $x \leq n$  e  $x \leq M \iff x \leq \min(n, M)$   
 $n - x \leq n$  e  $n - x \leq N - M \iff x \geq 0$  e  $x \geq n - (N - M)$   
 $\iff x \geq \max(0, n - (N - M))$

Então

$$S = \{\max(0, n - (N - M)), 1, \dots, \min(n, M)\}$$



- O número de casos favoráveis à saída de  $x$  sucessos, nas  $n$  extracções, é  $\binom{M}{x} \binom{N-M}{n-x}$
- O número de casos possíveis nas  $n$  extracções, é  $\binom{N}{n}$
- A probabilidade de obter exactamente  $x$  sucessos, nas  $n$  extracções, é dada por

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}, \text{ para } x \in S$$

$X$  é então uma v.a. discreta de Suporte  $S = \{\max(0, n - (N - M)), 1, \dots, \min(n, M)\}$  e função de probabilidade dada por

$$p(x) = P(X = x) = \begin{cases} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} & \text{se } x \in S \\ 0 & \text{se } x \in \bar{S} \end{cases}.$$

Diz-se que a v.a. **X segue a lei Hipergeométrica** (ou **X tem distribuição Hipergeométrica**) de parâmetros  $n$ ,  $N$  e  $M$ , e escreve-se simbolicamente  $\mathbf{X} \sim \mathcal{H}(n, N, M)$ .

### Propriedades

- Se  $X \sim \mathcal{H}(n, N, M)$  então

$$E(X) = n \frac{M}{N}$$

$$V(X) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \left(\frac{N-n}{N-1}\right)$$

### Notas

Tal como no caso da Binomial, a distribuição Hipergeométrica é usada na contagem do número de sucessos que ocorrem em  $n$  realizações de uma experiência aleatória, cada uma com apenas dois resultados possíveis (sucesso ou insucesso). No entanto, a Hipergeométrica é usada quando a probabilidade do sucesso é alterada em cada realização.

Uma das aplicações mais interessantes destas distribuições é na área da Amostragem, onde uma amostra de dimensão  $n$  é seleccionada numa população de dimensão  $N$ . Uma nova leitura dos parâmetros da Hipergeométrica,  $(n, N, M)$ , é

$n$ : dimensão da amostra

$N$ : dimensão da população (finita)

$M$ : número de sucessos na população

Prova-se, no entanto, que em certas condições, a distribuição  $\mathcal{H}(n, N, M)$  pode ser aproximada pela distribuição  $\mathcal{B}(n, p)$ , com  $p = \frac{M}{N}$ .

Na prática,

$$\text{se } \frac{n}{N} \leq 0.1 \text{ então } \mathcal{H}(n, N, M) \sim \mathcal{B}\left(n, \frac{M}{N}\right)$$

$\downarrow$

*“aproximadamente (no limite)”*

Intuitivamente, se a dimensão da amostra,  $n$ , é muito pequena relativamente à dimensão da população,  $N$ , a selecção, embora seja feita sem reposição, não vai alterar “significativamente” a probabilidade de ocorrer um sucesso.

(Mais à frente voltaremos a esta aproximação !!)

### 2.3.4 Distribuição de Poisson

Seja  $\lambda$  um número real positivo. Diz-se que uma v.a.  $X$  segue a lei de Poisson de parâmetro  $\lambda$ , e escreve-se simbolicamente  $X \sim \mathcal{P}(\lambda)$ , se  $X$  for discreta de suporte  $S = \mathbb{N}_0$ , com função de probabilidade

$$p(x) = P(X = x) = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!} & \text{se } x \in \mathbb{N}_0 \\ 0 & \text{se } x \in \overline{S} \end{cases}.$$

- $P(X = x) > 0, \forall x \in \mathbb{N}_0$

- $\sum_{x=0}^{+\infty} P(X = x) = 1$

#### Propriedades

- $E(X) = V(X) = \lambda$

- Sejam  $X_1, X_2, \dots, X_k$  v.a.'s independentes tais que  $X_i \sim \mathcal{P}(\lambda_i)$ . Então

$$Y = \sum_{i=1}^k X_i \sim \mathcal{P}(\lambda_1 + \lambda_2 + \dots + \lambda_k)$$

## Notas

Concluir que uma certa v.a. segue a lei de Poisson (ao contrário das leis Binomial e Hipergeométrica) não é imediato, e sai fora do âmbito do nosso curso. No entanto, esta distribuição é aplicada na contagem de eventos independentes que ocorrem durante um dado intervalo de tempo ou numa dada região espacial. Esta distribuição também é conhecida pela distribuição dos acontecimentos raros, pois a probabilidade de ocorrer mais do que um evento num intervalo muito pequeno é nula.

Prova-se que a distribuição  $\mathcal{B}(n, p)$ , com  $p$  muito pequeno e  $n$  suficientemente grande, pode ser aproximada pela distribuição  $\mathcal{P}(\lambda)$ , com  $\lambda = np$ . Na prática,

se  $p \leq 0.1$  então  $\mathcal{B}(n, p) \sim \mathcal{P}(np)$

## Exemplos

- 1- Considere um lote com 5 bolas, das quais 2 são brancas e 3 são vermelhas.

*Experiência aleatória:* Extração de uma bola, ao acaso, do lote.

Extração sucessiva, com reposição, de 3 bolas do lote.

$X$ : número de **bolas brancas** nas 3 extraídas

Qual é a probabilidade de saírem 2 bolas brancas?  $P(X = 2) = ?$

Qual é a probabilidade de saírem, no máximo, 2 bolas brancas?  $P(X \leq 2) = ?$

- 2- Um avião comercial tem 4 motores independentes e num voo a probabilidade de cada motor funcionar, sem avarias, é de 0.9. Qual a probabilidade do avião fazer uma viagem segura se para isso precisar de pelo menos 2 motores a funcionar correctamente?

$X$ : número de motores a funcionar sem avarias, em 4

$P(X \geq 2) = ?$

$Y = 4 - X$ : número de motores com avarias, em 4

$P(X \geq 2) = P(Y \leq 2) = ?$

- 3- Um material radioactivo emite um certo tipo de partículas a uma taxa média de duas por milissegundo, e segue uma distribuição de Poisson.

Qual é a probabilidade de:

serem emitidas 3 partículas num milissegundo?

serem emitidas 5 partículas em dois milissegundos?

serem emitidas pelo menos 3 partículas em dois milissegundos?

## 2.4 Variáveis Aleatórias Bidimensionais

**Definição** Seja  $\Omega$  o espaço de resultados associado a uma experiência aleatória. Designa-se por *variável aleatória bidimensional* (ou *vector aleatório de dimensão dois*) uma função  $\mathbf{X}$  de domínio  $\Omega$  e contradomínio em  $\mathbb{R}^2$ .

$$\mathbf{X} : \Omega \longrightarrow \mathbb{R}^2$$

$$w \longrightarrow \mathbf{X}(w) = (X_1(w), X_2(w))$$

### Exemplo de Ilustração

Lançamento de dois dados, anotando os números de pontos das faces que ficam voltadas para cima.

$$\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}.$$

Seja  $\mathbf{X} = (X_1, X_2)$  o vector aleatório que representa a **soma**,  $X_1$ , e a **diferença**,  $X_2$ , dos pontos obtidos.

### Notas

– Uma variável aleatória bidimensional é um vector cujas componentes são variáveis aleatórias unidimensionais.

– Associado a uma experiência aleatória, pode interessar o estudo de  $k$ , com  $k \geq 2$ , medidas/características. Surge o conceito de variável (vector) aleatória(o)  $k$ -dimensional, com

$$\mathbf{X}(w) = (X_1(w), X_2(w), \dots, X_k(w)) \in \mathbb{R}^k.$$

Por uma questão de clareza e simplicidade na exposição dos conceitos, no nosso estudo vamos considerar apenas variáveis aleatórias bidimensionais. No entanto, todos os conceitos aqui apresentados generalizam-se facilmente para o caso  $k$ -dimensional.

– Os vectores aleatórios podem ser discretos (todas as componentes do vector são v.a.'s discretas), contínuos (todas as componentes do vector são v.a.'s contínuas) ou mistos. Embora neste curso se analise apenas o caso discreto, os conceitos aqui apresentados generalizam-se para o caso contínuo. O caso misto está fora do âmbito do nosso estudo.

**Definição** Um vector aleatório  $(X, Y)$  diz-se discreto se existir um conjunto finito ou infinito numerável  $S$  tal que

$$S = \{(x, y) \in \mathbb{R}^2 : P(X = x, Y = y) > 0\}$$

e tal que

$$P((X, Y) \in S) = \sum_{(x, y) \in S} P(X = x, Y = y) = 1.$$

**Definição** Sejam  $X$  e  $Y$  duas variáveis aleatórias. A *função de probabilidade conjunta* do par  $(X, Y)$  é dada por

$$p_{XY}(x, y) = P(X = x \cap Y = y), \quad \forall (x, y) \in \mathbb{R}^2.$$

↓

“probabilidade de, simultaneamente,  
 $X$  tomar o valor  $x$  e  $Y$  tomar o valor  $y$ ”

### Propriedades

- (1)  $0 \leq p_{XY}(x, y) \leq 1, \forall (x, y) \in \mathbb{R}^2$
- (2)  $\sum_{(x, y) \in \mathbb{R}^2} p_{XY}(x, y) = \sum_{y \in \mathbb{R}} \sum_{x \in \mathbb{R}} p_{XY}(x, y) = 1.$

**Definição** A *função distribuição conjunta* do par de v.a.'s  $(X, Y)$  é dada por

$$F_{XY}(x, y) = P(X \leq x \cap Y \leq y), \quad \forall (x, y) \in \mathbb{R}^2.$$

Se  $(X, Y)$  tem suporte  $S = S_X \times S_Y$  então

$$\begin{aligned} F_{XY}(x, y) &= \sum_{(\mathbf{a}, \mathbf{b}) \in S \cap ]-\infty, x] \times ]-\infty, y]} P(X = \mathbf{a} \cap Y = \mathbf{b}). \\ &= \sum_{\mathbf{a} \in S_X \cap ]-\infty, x]} \sum_{\mathbf{b} \in S_Y \cap ]-\infty, y]} P(X = \mathbf{a} \cap Y = \mathbf{b}). \end{aligned}$$

### Algumas propriedades da função distribuição conjunta

- (1)  $\forall (x, y) \in \mathbb{R}^2, 0 \leq F_{XY}(x, y) \leq 1$  ( $F_{XY}$  é uma função limitada)
- (2)  $\forall (x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ , se  $x_1 < x_2$  e  $y_1 < y_2 \Rightarrow F_{XY}(x_1, y_1) \leq F_{XY}(x_2, y_2)$  ( $F_{XY}$  é não decrescente)
- (3)  $\lim_{(x,y) \rightarrow (-\infty, -\infty)} F_{XY}(x, y) = 0$  e  $\lim_{(x,y) \rightarrow (+\infty, +\infty)} F_{XY}(x, y) = 1$

**Definições** Seja  $(X, Y)$  um vector aleatório discreto com função de probabilidade conjunta  $p_{XY}$ .

- A *função de probabilidade marginal de X* é a função dada por

$$p_X(x) = P(X = x) = \sum_{y \in \mathbb{R}} p_{XY}(x, y) = \sum_{y \in \mathbb{R}} P(X = x \cap Y = y), \forall x \in \mathbb{R}.$$

- A *função de probabilidade marginal de Y* é a função dada por

$$p_Y(y) = P(Y = y) = \sum_{x \in \mathbb{R}} p_{XY}(x, y) = \sum_{x \in \mathbb{R}} P(X = x \cap Y = y), \forall y \in \mathbb{R}.$$

(As funções de probabilidade marginais são funções de probabilidade de v.a.'s unidimensionais.)

- A *função de probabilidade condicionada de Y dado X = x*, com  $p_X(x) > 0$ , é a função de  $y$  dada por

$$p_{Y/X=x}(y) = P(Y = y/X = x) = \frac{p_{XY}(x, y)}{p_X(x)}, \forall y \in \mathbb{R}.$$

(de modo análogo se define função de probabilidade condicionada de X dado Y = y, com  $p_Y(y) > 0$ )

$p_{Y/X=x}$  é uma função de probabilidade.

### Independência de Variáveis Aleatórias

**Definição** Sejam X e Y duas variáveis aleatórias discretas. X e Y dizem-se independentes se e só se

$$\forall (x, y) \in \mathbb{R}^2, p_{XY}(x, y) = p_X(x) \times p_Y(y).$$

**Generalização** As  $n$  ( $n \geq 2$ ) variáveis aleatórias  $X_1, X_2, \dots, X_n$  dizem-se independentes se e só se

$$\forall x_1, x_2, \dots, x_n \in \mathbb{R}, P(X_1 = x_1 \cap X_2 = x_2 \cap \dots \cap X_n = x_n) = \prod_{i=1}^n P(X_i = x_i).$$

## Parâmetros de Vectores Aleatórios

**Definição** Seja  $(X, Y)$  um vector aleatório tal que  $E(X)$  e  $E(Y)$  existem. Chama-se *esperança matemática* (valor médio ou valor esperado) do vector  $(X, Y)$  ao vector  $(E(X), E(Y))$ .

**Propriedade** Seja  $(X, Y)$  um vector aleatório discreto de suporte  $S$  e  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  tal que  $g(X, Y)$  é uma variável aleatória. Tem-se

$$E[g(X, Y)] = \sum_{(x, y) \in S} g(x, y) P(X = x, Y = y).$$

**Definição** Seja  $(X, Y)$  um vector aleatório discreto tal que  $E(X) = \mu_X$  e  $E(Y) = \mu_Y$ . A **Covariância entre  $X$  e  $Y$** , denotada por  $cov(X, Y)$ , é definida por

$$cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

$$\begin{aligned} \bullet \quad cov(X, Y) &= \sum_{x \in S_X} \sum_{y \in S_Y} (x - \mu_X)(y - \mu_Y) P(X = x, Y = y) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

**Algumas propriedades :**

1.  $V(X + Y) = V(X) + V(Y) + 2 cov(X, Y)$
2.  $V(X - Y) = V(X) + V(Y) - 2 cov(X, Y)$
3. Se  $X_1, X_2, \dots, X_n$  ( $n \in \mathbb{N}$ ) são v.a.'s **independentes** então

$$E(X_1 X_2 \dots X_n) = E(X_1)E(X_2) \dots E(X_n);$$

$$V\left[\sum_{i=1}^n a_i X_i + b\right] = \sum_{i=1}^n a_i^2 V(X_i), \quad \forall a_1, a_2, \dots, a_n, b \in \mathbb{R}.$$

Casos particulares:

- i)  $V\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n V(X_i);$
- ii)  $V(X_1 - X_2) = V(X_1) + V(X_2).$

**Nota** Se  $X$  e  $Y$  são independentes então  $cov(X, Y) = 0$  (!), mas a implicação contrária já não é válida!

**Definição** Seja  $(X, Y)$  um vector aleatório. O **coeficiente de correlação linear entre  $X$  e  $Y$** , denotado por  $\rho_{XY}$ , é dado por

$$\rho_{XY} = \frac{cov(X, Y)}{\sqrt{V(X) V(Y)}} = \frac{cov(X, Y)}{\sigma_X \sigma_Y}.$$



**Nota**

$$-1 \leq \rho_{XY} \leq 1$$

**Exemplo**

- 1- Uma fábrica possui duas linhas de produção de um certo motor. Considere o vector aleatório  $(X, Y)$ , onde

$X$ : número de motores produzidos, diariamente, na linha I

$Y$ : número de motores produzidos, diariamente, na linha II

A função de probabilidade conjunta de  $(X, Y)$  é dada na forma tabular, por

X	0	1	2
Y			
0	0	0.06	0.19
1	0.03	0.07	0.18
2	0.04	0.06	0.1
3	0.04	0.08	0.15

Funções de probabilidade marginais

x	0	1	2
$p_X(x)$	0.11	0.27	0.62

y	0	1	2	3
$p_Y(y)$	0.25	0.28	0.2	0.27

$$E(X) = 1.51, \quad V(X) = 0.4699, \quad \sigma_X = 0.6855$$

$$E(Y) = 1.49, \quad V(Y) = 1.2899, \quad \sigma_Y = 1.1357$$

(Comparar as duas linhas de produção.)

## 3 Variáveis Aleatórias e Distribuições de Probabilidade Contínuas

### 3.1 Variáveis Aleatórias Contínuas

**Definição** Uma v.a.  $X$  diz-se contínua se existir uma função  $f : \mathbb{R} \rightarrow \mathbb{R}$  que verifique

- (i)  $f(x) \geq 0, \forall x \in \mathbb{R}$  (não negativa)
- (ii)  $\int_{-\infty}^{+\infty} |f(x)| dx < \infty$  (integrável em  $\mathbb{R}$ )
- (iii)  $\int_{-\infty}^{+\infty} f(x) dx = 1$

e tal que  $\forall a, b \in \mathbb{R}, P(X \in ]a, b[) = \int_a^b f(x) dx$ .

Uma função  $f$  com estas características chama-se **função densidade de probabilidade** de  $X$ , e permite calcular a probabilidade de  $X$  assumir valores num qualquer intervalo real.

**Observações** Se  $X$  é uma v.a. contínua com densidade  $f$ ,

$$\forall a \in \mathbb{R}, P(X = a) = \int_a^a f(x) dx = 0$$

$$\forall a, b \in \mathbb{R}, P(a \leq X \leq b) = P(a < X < b) = \int_a^b f(x) dx$$

$$\forall a \in \mathbb{R}, P(X \geq a) = P(X > a) = \int_a^{+\infty} f(x) dx$$

**Definição** Seja  $X$  uma v.a. contínua com função densidade de probabilidade  $f$ . Chama-se **função distribuição de  $X$**  à função  $F : \mathbb{R} \rightarrow [0, 1]$  definida por

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \int_{-\infty}^x f(a) da, \quad \forall x \in \mathbb{R}. \end{aligned}$$

**Nota** A definição de função distribuição é a mesma quer as variáveis aleatórias sejam discretas ou contínuas (a integração no caso contínuo é a extensão “natural” da soma (caso discreto)).

#### Parâmetros de localização e de dispersão

Os parâmetros de localização e de dispersão de uma v.a. contínua definem-se de modo análogo ao caso discreto.

#### Definições

Seja  $X$  uma v.a. contínua de densidade  $f$ .

- A *esperança matemática de  $X$* , caso exista, é definida por

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx.$$

(A existência do valor médio de  $X$  depende da convergência do integral anterior)

▷ Seja  $g : \mathbb{R} \rightarrow \mathbb{R}$  tal que  $g(X)$  é uma v.a. contínua. Então

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) f(x) dx.$$

Exemplo:  $k \in \mathbb{N}$  (arbitrariamente fixo),  $E(X^k) = \int_{-\infty}^{+\infty} x^k f(x) dx$

- Considere  $E(X) = \mu$ . A *variância de  $X$*  é definida por

$$V(X) = E[(X - \mu)^2] = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx.$$

- A *mediana de  $X$*  é o número real  $M_d$  tal que

$$F(M_d) = 0.5.$$

(Note que  $F$  é uma função contínua)

## 3.2 Distribuições Especiais Contínuas

### 3.2.1 Distribuição Uniforme (Contínua)

Seja  $[a, b]$  um intervalo real não vazio. Diz-se que uma v.a. **X segue a lei** (ou **X tem distribuição**) **Uniforme no intervalo**  $[a, b]$ , e escreve-se simbolicamente  $\mathbf{X} \sim \mathcal{U}_{[a,b]}$ , se  $X$  for uma variável contínua com função densidade

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } x \in [a, b] \\ 0 & \text{se } x < a \vee x > b \end{cases} = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x)$$

A função distribuição de  $X$  é dada por

$$F(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{b-a} & \text{se } x \in [a, b] \\ 1 & \text{se } x > b \end{cases}.$$

#### Propriedades

- Se  $X \sim \mathcal{U}_{[a,b]}$  então

$$E(X) = \frac{a+b}{2}$$

$$V(X) = \frac{(b-a)^2}{12}$$

#### Nota

A distribuição Uniforme é usada para descrever medidas que variam aleatoriamente num certo intervalo não vazio  $[a, b]$  e cuja probabilidade de tomar valores num sub-intervalo de  $[a, b]$  é proporcional ao seu comprimento. De facto, se  $X \sim \mathcal{U}_{[a,b]}$ ,

$$\forall [c, d] \subseteq [a, b], P(c \leq X \leq d) = k(d - c), \text{ com } k = \frac{1}{b-a}.$$

### 3.2.2 Distribuição Exponencial

Seja  $\lambda$  um parâmetro real positivo. Diz-se que uma v.a. **X tem distribuição Exponencial de parâmetro**  $\lambda$ , e escreve-se simbolicamente  $\mathbf{X} \sim \mathcal{E}(\lambda)$ , se  $X$  for uma variável contínua com função densidade

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases} = \lambda e^{-\lambda x} \mathbf{1}_{[0,+\infty)}(x)$$

A função distribuição de  $X$  é dada por

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 - e^{-\lambda x} & \text{se } x \geq 0 \end{cases}.$$

### Propriedades

- Se  $X \sim \mathcal{E}(\lambda)$  então

$$E(X) = \frac{1}{\lambda}$$

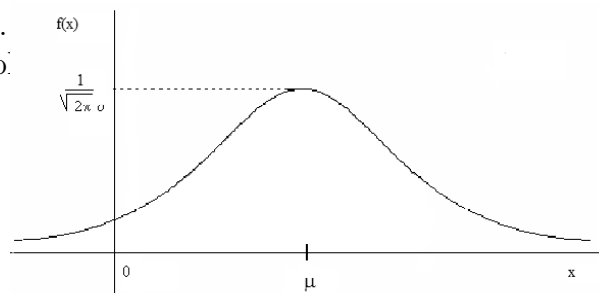
$$V(X) = \frac{1}{\lambda^2}$$

### Nota

Se a distribuição de Poisson é aplicada na contagem de eventos independentes num certo intervalo de tempo, a distribuição Exponencial é usada para representar intervalos de tempo entre eventos independentes!!

### 3.2.3 Distribuição Normal

Sejam  $\mu \in \mathbb{R}$  e  $\sigma \in \mathbb{R}^+$ .  
 $\mu$  e  $\sigma$ , e escreve-se símbolo  
densidade



**Normal de parâmetros**  
ável contínua com função

**Nota** Diz-se também que  $X$  é Normalmente distribuída, ou que  $X$  segue a lei Normal, ou ainda que  $X$  é uma v.a. Gaussiana (de Gauss), de parâmetros  $\mu$  e  $\sigma$ .

**Representação gráfica de  $f$  ( $\mu > 0$ )**

### Características principais:

- Forma de “sino”
- Máximo global para  $x = \mu$
- Simétrica relativamente a  $x = \mu$

- $\lim_{x \rightarrow -\infty} f(x) = \lim_{x \rightarrow +\infty} f(x) = 0$  (eixo  $xx$  é uma assíntota)

Para  $X \sim \mathcal{N}(\mu, \sigma)$ , de densidade  $f$ ,

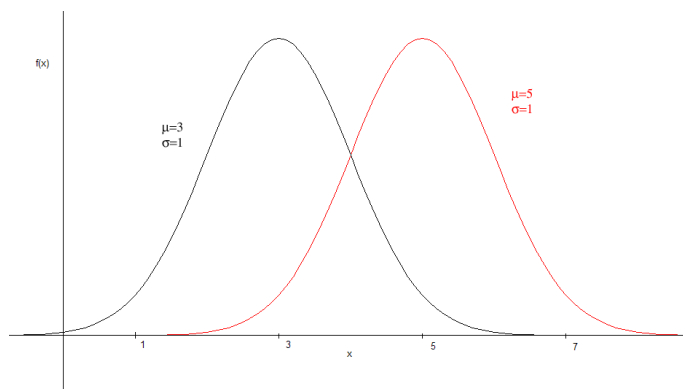
- $\int_{-\infty}^{+\infty} f(x) dx = 1$
- $P(X \in [a, b]) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \dots!?$
- $P(X > \mu) = P(X < \mu) = 0.5$ .

### Parâmetros

Mostra-se que

$$E(X) = \mu \quad (\text{parâmetro de localização})$$

$$V(X) = \sigma^2 \implies \sigma(X) = \sigma \quad (\text{parâmetro de dispersão})$$



**Caso particular:**  $\mu = 0$  e  $\sigma = 1$ .

- A distribuição  $\mathcal{N}(0, 1)$  é chamada normal estandardizada (*standard*) ou normal centrada e reduzida.
- Se  $Z \sim \mathcal{N}(0, 1)$  então

$$F_Z(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt, \forall z \in \mathbb{R}.$$

- Estão disponíveis (em tabelas ou máquinas de calcular) as probabilidades  $F_Z(z)$ , para alguns valores de  $z$ .

Em particular, nas tabelas de PE podem ser consultadas as probabilidades  $F_Z(z)$ , para alguns  $z \geq 0$ !



### Propriedades da Lei Normal

1. Se  $\mathbf{X} \sim \mathcal{N}(\mu, \sigma)$  então  $\mathbf{Z} = \frac{\mathbf{X} - \mu}{\sigma} \sim \mathcal{N}(\mathbf{0}, 1)$ .

$$\begin{aligned} \bullet P(a < X < b) &= P\left(\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right) \\ &= P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right) \\ &= F_Z\left(\frac{b - \mu}{\sigma}\right) - F_Z\left(\frac{a - \mu}{\sigma}\right) \end{aligned}$$

2. Se  $X \sim \mathcal{N}(\mu, \sigma)$  e  $a, b \in \mathbb{R}$  ( $a \neq 0$ ) então

$$Y = aX + b \sim \mathcal{N}(a\mu + b, \sqrt{a^2\sigma^2})$$

### 3. (Estabilidade da Lei Normal)

Sejam  $X_1, X_2, \dots, X_n$  v.a.'s independentes e tais que  $X_i \sim \mathcal{N}(\mu_i, \sigma_i)$ , com  $\mu_i \in \mathbb{R}$  e  $\sigma_i \in \mathbb{R}^+$  para  $i = 1, 2, \dots, n$ . Então

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

“a combinação linear de v.a.'s independentes com distribuição normal é (ainda) uma v.a. normalmente distribuída”

#### Casos particulares

Sejam  $X_1, X_2, \dots, X_n$  v.a.'s independentes e tais que  $X_i \sim \mathcal{N}(\mu, \sigma)$ , com  $\mu \in \mathbb{R}$  e  $\sigma \in \mathbb{R}^+$  para  $i = 1, 2, \dots, n$ . Então

$$\begin{aligned} \sum_{i=1}^n X_i &\sim \mathcal{N}(n\mu, \sqrt{n}\sigma) \\ \frac{1}{n} \sum_{i=1}^n X_i &\sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \end{aligned}$$

#### Aproximações da Binomial e da Poisson à Normal

As aproximações das distribuições Binomial e Poisson à distribuição normal são consequência do Teorema do Limite Central (Capítulo 4).

- Se  $X \sim \mathcal{B}(n, p)$  [ $E(X) = np$  e  $V(X) = np(1-p)$ ], para  $n$  *suficientemente grande* e  $p \in ]0.1, 0.9[$ ,

$$X \dot{\sim} \mathcal{N}(np, \sqrt{np(1-p)}) \Leftrightarrow Z = \frac{X - np}{\sqrt{np(1-p)}} \dot{\sim} \mathcal{N}(0, 1)$$

- Se  $X \sim \mathcal{P}(\lambda)$  [ $E(X) = V(X) = \lambda$ ], para  $\lambda$  *suficientemente grande*,

$$X \dot{\sim} \mathcal{N}(\lambda, \sqrt{\lambda}) \Rightarrow Z = \frac{X - \lambda}{\sqrt{\lambda}} \dot{\sim} \mathcal{N}(0, 1)$$

#### Notas

- A aproximação à Normal é tanto melhor quanto maior for o valor de  $n$  (no caso da Binomial) ou de  $\lambda$  (caso da Poisson), mas na prática pode ser efectuada com resultados razoáveis desde que  $n > 20$  ou  $\lambda > 20$ .

- Na aproximação de uma distribuição discreta à normal, como nos casos anteriores, deve estar presente que a normal é uma distribuição contínua. Com o objectivo de reduzir o erro de aproximação é usual efectuar-se uma **correção de continuidade**. Não a efectuaremos nas aulas de PE, mas fica a nota de que se  $X$  é uma v.a. discreta, a **correção de continuidade** consiste em *converter*  $X$  numa v.a. contínua, do modo seguinte (por exemplo):

$$P(X = x) \simeq P(x - 0.5 \leq X \leq x + 0.5)$$



### 3.2.4 Distribuição Qui-Quadrado

Seja  $n \in \mathbb{N}$ . Diz-se que uma v.a. **X tem distribuição Qui-quadrado com n graus de liberdade**, e escreve-se simbolicamente  $\mathbf{X} \sim \chi_n^2$ , se  $X$  for uma variável contínua com função densidade

$$f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} e^{-\frac{x}{2}} x^{\frac{n}{2}-1}, \quad \forall x \in \mathbb{R}^+,$$

onde  $\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$ ,  $\alpha > 0$  (chamada função Gama)

• Se  $Z_1, Z_2, \dots, Z_n$  são v.a.'s independentes e identicamente distribuídas com  $Z_i \sim \mathcal{N}(0, 1)$ , então  $\sum_{i=1}^n Z_i^2 \sim \chi_n^2$ .

• Se  $X \sim \chi_n^2$  então 
$$\begin{aligned} E(X) &= n \\ V(X) &= 2n \end{aligned}$$

#### Notas

- A distribuição Qui-quadrado é uma distribuição **assimétrica**, enviesada à direita. À medida que o número de graus de liberdade  $n$  aumenta, o enviesamento diminui e aproxima-se da distribuição normal.
- A distribuição Qui-quadrado tem um papel fundamental *no estudo* da variância de uma população, como veremos mais tarde.
- **Tabela de PE :**

Seja  $X \sim \chi_n^2$ .

Para alguns valores de  $n \in \mathbb{N}$  e  $p \in ]0, 1[$ , pode ser consultado na Tabela de PE o quantil de ordem  $p$  de  $X$ , isto é,  $x_p$  :

$$P(X \leq x_p) = p \Leftrightarrow F_X(x_p) = p \Leftrightarrow x_p = F_X^{-1}(p).$$

### 3.2.5 Distribuição t-Student

Seja  $n \in \mathbb{N}$ . Diz-se que uma v.a. **X tem distribuição t-Student com n graus de liberdade**, e escreve-se simbolicamente  $\mathbf{X} \sim \mathbf{t}_n$ , se  $X$  for uma variável contínua com função densidade

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad \forall x \in \mathbb{R},$$

com  $\Gamma$  a função Gama.

- Se  $Z$  e  $Y$  são v.a.'s independentes tais que  $Z \sim \mathcal{N}(0, 1)$  e  $Y \sim \chi_n^2$ , então

$$\frac{Z}{\sqrt{\frac{Y}{n}}} \sim t_n.$$

- Se  $X \sim t_n$  então  
 $E(X) = 0$  se  $n > 1$  (**centrada**; para  $n = 1$  não existe esperança de  $X$ )

$$V(X) = \frac{n}{n-2} \text{ se } n > 2$$

## Notas

- A distribuição t-Student é uma distribuição **simétrica** relativamente a  $x = 0$  e tem forma de “sino”, tal como a normal, mas a t-Student tem caudas mais “pesadas”. À medida que o número de graus de liberdade  $n$  aumenta ( $n > 30$ ), aproxima-se da distribuição normal centrada e reduzida.
- A distribuição t-Student tem um papel fundamental *no estudo* da esperança de uma população.
- **Tabela de PE** : Para  $X \sim t_n$ , pode ser consultado o quantil de ordem  $p$  de  $X$ , para alguns valores de  $n \in \mathbb{N}$  e  $p \in ]0, 1[$ .

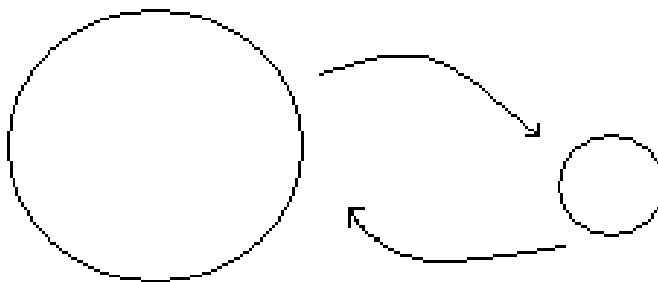
## 4 Amostragem e Distribuições Amostrais

Neste capítulo serão dadas algumas noções e resultados teóricos fundamentais em Estatística. O estudo de características populacionais, a partir de amostras, assenta nestes resultados.

### 4.1 Amostra Aleatória. Estatísticas

População  $X$

Amostra Aleatória  $X_1, X_2, \dots, X_n$



#### Definições

Uma **amostra aleatória** da v.a.  $X$  é um conjunto de  $n$  v.a.'s  $X_1, X_2, \dots, X_n$  independentes e com a mesma distribuição de  $X$  (dizem-se independentes e identicamente distribuídas (i.i.d.) com  $X$ ).

**Estatística** é uma função da amostra aleatória, e logo uma variável aleatória, cuja expressão não contém parâmetros desconhecidos.

Sejam  $X_1, X_2, \dots, X_n$  uma amostra aleatória de  $X$ .

- A **média da amostra** ou **média amostral**, denotada por  $\bar{X}$ , é dada por

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

- A **variância da amostra** ou **variância amostral**, denotada por  $S_n^2$ , é dada por

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

- O **desvio padrão amostral** é dado por  $S_n = \sqrt{S_n^2}$ .

**Exercício:** Verifique que  $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n X_i^2 - \frac{n}{n-1} \bar{X}^2$ .

A média e a variância amostrais são estatísticas (especiais no nosso estudo) e, logo, variáveis aleatórias, com distribuições próprias, com parâmetros de localização e dispersão, que dependem da população em estudo. As suas distribuições são designadas *distribuições amostrais*.

## 4.2 Distribuição da Média Amostral

Considere-se uma amostra aleatória  $X_1, X_2, \dots, X_n$  da v.a. (população)  $X$  com parâmetros  $E(X) = \mu$  e  $V(X) = \sigma^2$ .

Usando propriedades da esperança e da variância, verifica-se que:

$$E(\bar{X}) = \mu \quad \text{e} \quad V(\bar{X}) = \frac{\sigma^2}{n} \Rightarrow \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

### Teorema do Limite Central (T.L.C.)

Sejam  $X_1, X_2, \dots, X_n$  uma amostra aleatória da v.a. (população)  $X$  com parâmetros  $E(X) = \mu$  e  $V(X) = \sigma^2$ . Para  $n$  *suficientemente* grande, tem-se

$$\frac{\bar{X} - E(\bar{X})}{\sigma(\bar{X})} = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

### Notas

- A aproximação à Normal centrada e reduzida pode ser efectuada na prática para  $n > 30$ ; no entanto, esta será tanto melhor quanto maior for o valor de  $n$ .
- Se a população é normal, isto é,  $X \sim \mathcal{N}(\mu, \sigma)$  (caso já estudado) então

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \Rightarrow \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \sim \mathcal{N}(0, 1) \text{ (distribuição exacta!)}$$

O T.L.C. permite alargar este resultado a qualquer família, desde que se considere uma amostra de tamanho bastante razoável!

### Justificação de alguns resultados apresentados anteriormente

- Se  $X \sim \mathcal{B}(n, p)$  [ $E(X) = np$  e  $V(X) = np(1-p)$ ], para  $n$  *suficientemente grande* e  $p \in ]0.1, 0.9[$ ,

$$X \sim \mathcal{N}(np, \sqrt{np(1-p)}) \Leftrightarrow Z = \frac{X - np}{\sqrt{np(1-p)}} \sim \mathcal{N}(0, 1)$$

### Nota

Neste caso,  $X = \sum_{i=1}^n X_i$  com  $X_i \sim \mathcal{B}(p)$  para  $i = 1, \dots, n$ .

**Exercício:** Deduza que

$$Z_n = \sqrt{n} \frac{\bar{X} - p}{\sqrt{p(1-p)}} \sim \mathcal{N}(0, 1)$$

Anteriormente, vimos que

se  $p \leq 0.1$  então  $X \sim \mathcal{P}(np)$ ;

se  $p \geq 0.9$  então  $Y = n - X \sim \mathcal{P}(n(1-p))$

- Se  $X \sim \mathcal{P}(\lambda)$  [ $E(X) = V(X) = \lambda$ ], para  $\lambda$  *suficientemente grande*,

$$X \dot{\sim} \mathcal{N}(\lambda, \sqrt{\lambda}) \Rightarrow Z = \frac{X - \lambda}{\sqrt{\lambda}} \dot{\sim} \mathcal{N}(0, 1)$$

**Nota**

Neste caso,  $X = \sum_{i=1}^n X_i$  com  $X_i \sim \mathcal{P}(\frac{\lambda}{n})$  para  $i = 1, \dots, n$ .

**Distribuições da média amostral  $\bar{X}$ :**

- Para  $n$  *suficientemente grande*,  $\sqrt{n} \frac{\bar{X} - \mu}{\sigma} \dot{\sim} \mathcal{N}(0, 1)$ .
- Se  $X \sim \mathcal{N}(\mu, \sigma)$ ,

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

$$\sqrt{n} \frac{\bar{X} - \mu}{S_n} \sim t_{n-1}$$

### 4.3 Distribuição da Variância Amostral

- Se  $X \sim \mathcal{N}(\mu, \sigma)$ ,  $\frac{n-1}{\sigma^2} S_n^2 \sim \chi_{n-1}^2$

## 5 Estimação

### 5.1 Introdução

#### Problema

Seja  $X$  a v.a. em estudo, com função de probabilidade/densidade  $f(x, \theta)$ , onde  $\theta$  é um (ou mais) parâmetro(s) desconhecido(s).

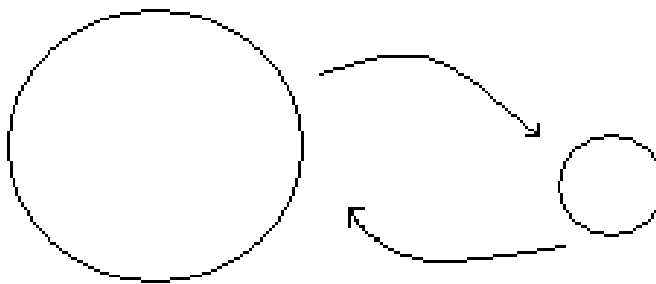
**Que valor(es) atribuir a  $\theta$  ? Ou a  $g(\theta)$  ?**

#### Exemplos

- $X \sim \mathcal{P}(\lambda)$        $f(x, \lambda) = e^{-\lambda} \frac{\lambda^x}{x!}, \forall x \in \mathbb{N}_0$        $\longrightarrow \theta = \lambda$
- $X \sim \mathcal{E}(\lambda)$        $f(x, \lambda) = \lambda e^{-\lambda x} \mathbf{1}_{[0, +\infty[}(x)$        $\longrightarrow \theta = \lambda, g(\lambda) = \frac{1}{\lambda}$
- $X \sim \mathcal{N}(\mu, \sigma)$        $f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \forall x \in \mathbb{R}$        $\longrightarrow \theta = (\mu, \sigma)$

População  $X$

Amostra Aleatória  $X_1, X_2, \dots, X_n$



*Inferência Estatística*

**Parâmetros Popacionais**

$$E(X) = \mu$$

$$V(X) = \sigma^2, \sigma(X) = \sigma, \dots$$

**Características Amostrais**

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

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, S_n, \dots$$

### 5.2 Estimação Pontual

#### Definições

**Estimador** é qualquer Estatística usada para estimar um parâmetro populacional (ou função desse parâmetro).

**Estimativa** é um valor do estimador para uma amostra em concreto.

**Notação** Estimador do parâmetro  $\theta \rightarrow \hat{\Theta}$

## Exemplos

População $X \longrightarrow$	a.a. $X_1, X_2, \dots, X_n \longrightarrow$	concretização da a.a. $x_1, x_2, \dots, x_n$
$\downarrow$	$\downarrow$	$\downarrow$
$\mu$	$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
$\sigma^2$	$\hat{\sigma}^2 = S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$	$s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
$\vdots$		
(parâmetros desconhecidos)	(Estimador)	(Estimativa)

### Observação

Os valores possíveis de um Estimador (note-se v.a.) são as estimativas obtidas em cada concretização da amostra aleatória.

### Algumas propriedades de um (*bom*) Estimador:

- Um estimador  $\hat{\Theta}$  do parâmetro  $\theta$  diz-se **cêntrico** (ou **centrado** ou **não enviesado**) se  $E(\hat{\Theta}) = \theta$ .
- Um estimador  $\hat{\Theta}$  do parâmetro  $\theta$  diz-se **enviesado** se não for cêntrico e o seu **enviesamento** é dado por  $viés(\hat{\Theta}) = E(\hat{\Theta}) - \theta$ .
- A **eficiência** de um estimador cêntrico  $\hat{\Theta}$  é dada por  $V(\hat{\Theta})$ .

Assim, dados dois estimadores cêntricos do parâmetro  $\theta$ ,  $\hat{\Theta}_1$  e  $\hat{\Theta}_2$ ,  $\hat{\Theta}_1$  diz-se **mais eficiente** do que  $\hat{\Theta}_2$  se  $V(\hat{\Theta}_1) < V(\hat{\Theta}_2)$ .

Entre estimadores não enviesados, é dada preferência ao estimador com menor variância, isto é, ao mais eficiente.

### Exemplo

Considerem-se duas a.a.'s de uma população  $X$ , de tamanhos  $n$  e  $m$ , respectivamente, com  $n > m$ . As médias amostrais  $\bar{X}_n$  e  $\bar{X}_m$  são estimadores cêntricos de  $\mu = E(X)$ ; no entanto,  $\bar{X}_n$  é mais eficiente do que  $\bar{X}_m$ . De facto,

$$V(\bar{X}_n) = \frac{V(X)}{n} < V(\bar{X}_m) = \frac{V(X)}{m}$$

- A **eficiência** de um estimador  $\hat{\Theta}$  do parâmetro  $\theta$  é dada pelo seu **erro quadrático médio** (**EQM**), definido por

$$EQM(\hat{\Theta}) = E[(\hat{\Theta} - \theta)^2] = V(\hat{\Theta}) + [viés(\hat{\Theta})]^2.$$

Entre estimadores enviesados e não enviesados, é dada preferência ao estimador com menor erro quadrático médio (o mais eficiente). Note-se que um estimador  $\hat{\Theta}_1$  enviesado pode ser mais eficiente do que um estimador cêntrico  $\hat{\Theta}_2$ , bastando que  $EQM(\hat{\Theta}_1) < V(\hat{\Theta}_2)$ .

### 5.3 Estimação por Intervalos

Como determinar um intervalo no qual se espera encontrar, com uma certa *confiança*, o valor de um parâmetro populacional  $\theta$  (ou de  $g(\theta)$ )?

#### Definição

Um intervalo de confiança  $(1 - \alpha)\%$  para o parâmetro  $\theta$  (ou  $g(\theta)$ ) é um intervalo aleatório  $]L_1, L_2[$ , com  $L_1$  e  $L_2$  duas estatísticas amostrais (v.a.'s) tais que

$$P(L_1 < \theta < L_2) = 1 - \alpha.$$

$$(\theta \in ]L_1, L_2[ \text{ com probabilidade } 1 - \alpha)$$

- $1 - \alpha$  é a probabilidade do intervalo conter  $\theta$  (o grau de confiança atribuído ao intervalo);
- $\alpha \in ]0, 1[$  é a probabilidade do intervalo não conter  $\theta$  (valor preferencialmente *pequeno*).

Como calcular  $L_1$  e  $L_2$ ?

#### Método da Variável Fulcral

Seja  $X$  uma v.a. cuja distribuição contém um parâmetro  $\theta$  desconhecido. Pretende-se um intervalo de confiança  $(1 - \alpha)\%$  para  $\theta$  (ou para  $g(\theta)$ ).

Considere-se uma amostra aleatória  $X_1, X_2, \dots, X_n$ , de  $X$ .

Nota: Variável Fulcral é uma função da amostra aleatória  $X_1, X_2, \dots, X_n$  e do parâmetro  $\theta$ , mas com distribuição (exacta ou aproximada) independente de  $\theta$ .

1- Escolher a Variável Fulcral  $Z_n$  (tabelas da disciplina)

2- Determinar  $z_1, z_2 \in \mathbb{R}$ :  $P(z_1 < Z_n < z_2) = 1 - \alpha$

$$\text{Nota: } z_1 : P(Z_n < z_1) = \frac{\alpha}{2} \quad \wedge \quad z_2 : P(Z_n < z_2) = 1 - \frac{\alpha}{2}$$

3- Com  $z_1$  e  $z_2$  conhecidos, encontrar

$$L_1 \equiv L_1(X_1, X_2, \dots, X_n) \text{ e } L_2 \equiv L_2(X_1, X_2, \dots, X_n)$$

tais que

$$P(z_1 < Z_n < z_2) = 1 - \alpha \Leftrightarrow P(L_1 < \theta < L_2) = 1 - \alpha$$

$$(\theta \in ]L_1, L_2[ \text{ com probabilidade } 1 - \alpha)$$

4- Para uma amostra particular  $x_1, x_2, \dots, x_n$ , determinar estimativas para  $L_1$  e  $L_2$ , respectivamente,

$$l_1 \equiv l_1(x_1, x_2, \dots, x_n) \text{ e } l_2 \equiv l_2(x_1, x_2, \dots, x_n)$$

Obtém-se, assim, uma estimativa para o intervalo de  $\theta$ , calculado com confiança  $(1 - \alpha)\%$ ,

$$IC_\theta = ]l_1, l_2[$$

**Nota:** O intervalo  $]L_1, L_2[$  é aleatório, com probabilidade  $1 - \alpha$  de conter  $\theta$ ; isto é, se forem recolhidas amostras em número bastante elevado, espera-se que  $(1 - \alpha)\%$  dos intervalos estimados com base nessas amostras contenha  $\theta$ ; o intervalo  $]l_1, l_2[$  é apenas uma estimativa, calculado com confiança  $(1 - \alpha)\%$ , mas ao qual não se pode atribuir uma “probabilidade” de conter  $\theta$ .



## Exemplo

### • Intervalo de confiança para o valor médio (caso de uma população normal com $\sigma$ conhecido)

Seja  $(X_1, X_2, \dots, X_n)$  uma a.a. relativa à v.a.  $X \sim \mathcal{N}(\mu, \sigma)$ , sendo  $\sigma$  conhecido. Vimos que um estimador para a média da população  $\mu$  é dado pela estatística  $\bar{X}$ . Neste caso, segue-se que

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

Fixado o valor de  $\alpha$  e designando por  $z_{\alpha/2}$  o valor tal que  $P(Z > z_{\alpha/2}) = \alpha/2$ , tem-se

$$\begin{aligned} P(-z_{\alpha/2} < Z < z_{\alpha/2}) &= 1 - \alpha &\Leftrightarrow & P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha \\ &&\Leftrightarrow & P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \end{aligned}$$

Para uma realização da a.a., ou seja, uma a.a. concreta  $(x_1, x_2, \dots, x_n)$ , e designando por  $\bar{x}$  o valor concreto da estatística  $\bar{X}$ , tem-se que

**o intervalo com  $(1 - \alpha) \times 100\%$  de confiança para o parâmetro  $\mu$ , numa população normal com  $\sigma$  conhecido, é:**

$$\mu \in \left[ \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

## Exercícios complementares

**1-** A resistência das cordas produzidas por determinada fábrica tem distribuição normal. A fábrica testou uma amostra de 51 cordas tendo obtido uma resistência média de 300 *Kg* e desvio padrão 24 *Kg*.

- (a) Determine um intervalo de confiança 95% para a resistência média deste tipo de cordas.
- (b) Determine um intervalo de confiança 90% para o desvio padrão.

**2-** O conteúdo médio de nicotina de uma amostra de 10 cigarros de uma certa marca é de 1 *mg* (miligramas). O laboratório sabe, pela longa experiência neste tipo de análise, que o conteúdo de nicotina é uma v.a. com distribuição normal de desvio padrão 0.15 *mg*.

- (a) Determine um intervalo de confiança 90% para o conteúdo médio de nicotina.
- (b) Com a mesma confiança, quantos cigarros deveriam ser analisados de modo a que o erro máximo cometido na estimação não ultrapasse 0.01?

**3-** Realizou-se um inquérito telefónico para estimar a proporção da população de um país, favorável a determinada reforma fiscal.

No inquérito realizado a 50 pessoas, escolhidas aleatoriamente, 84% manifestou-se favorável à reforma fiscal.

Determine um intervalo de confiança 95% para a proporção da população favorável à reforma fiscal.

## 6 Testes de Hipóteses Paramétricos

Vamos considerar neste capítulo outro aspecto do processo de inferência estatística: os **testes de hipóteses**. O objectivo é decidir se uma conjectura sobre determinada característica da(s) população(ões) em estudo é ou não apoiada pela evidência obtida de dados amostrais. Por exemplo, será que a média populacional difere de um valor pré-especificado?

As hipóteses serão formuladas sobre o(s) valor(es) do(s) parâmetro(s) desconhecidos da população,  $\theta$ . Nos testes paramétricos admitem-se certos pressupostos, uma lei de probabilidade para a população em estudo por exemplo, ao contrário do que acontece nos testes não paramétricos e que não serão abordados na disciplina.

### NOTAÇÃO E METODOLOGIA

O primeiro passo na condução de um teste de hipóteses é especificar as duas hipóteses do teste. Em geral, começa-se por formular a **hipótese alternativa** ( $H_1$ ), que é a hipótese que se pretende verificar (hipótese proposta pelo investigador).  $H_1$  contém, regra geral, uma desigualdade ( $<$ ,  $>$  ou  $\neq$ ). Definida  $H_1$ , formula-se a hipótese complementar de  $H_1$ , que se designa por **hipótese nula** ( $H_0$ ), e que é a hipótese que, à partida se julga inverosímil.  $H_0$  contém, regra geral, uma igualdade ( $=$ ).

Consideramos três tipos de testes:

1. Teste bilateral

$$\begin{aligned} H_0 : & \quad \theta = \theta_0 \\ H_1 : & \quad \theta \neq \theta_0 \end{aligned}$$

2. Teste unilateral à direita

$$\begin{aligned} H_0 : & \quad \theta = \theta_0 \quad (\text{ou } \theta \leq \theta_0) \\ H_1 : & \quad \theta > \theta_0 \end{aligned}$$

3. Teste unilateral à esquerda

$$\begin{aligned} H_0 : & \quad \theta = \theta_0 \quad (\text{ou } \theta \geq \theta_0) \\ H_1 : & \quad \theta < \theta_0 \end{aligned}$$

Dir-se-á que se pretende testar  $H_0$  contra  $H_1$ , ao que frequentemente se escreve:

$$\begin{aligned} & H_0 \text{ vs } H_1 \\ & \text{o réu é inocente } (H_0) \text{ vs o réu é culpado } (H_1) \end{aligned}$$

Ao efectuar um teste, assumimos sempre que a hipótese  $H_0$  é verdadeira, e não havendo evidência para provar o contrário decide-se não rejeitar  $H_0$ . Há duas possibilidades de se tomar uma decisão errada. Os dois **tipos de erros** que podem ser cometidos são:

**erro tipo I:** rejeitar  $H_0$  sendo  $H_0$  verdadeira;

**erro tipo II:** não rejeitar  $H_0$  sendo  $H_0$  falsa;

$$\begin{aligned} \alpha &= P(\text{erro tipo I}) \text{ (probabilidade de cometer um erro tipo I)} \\ \beta &= P(\text{erro tipo II}) \text{ (probabilidade de cometer um erro tipo II)} \end{aligned}$$

$\alpha$  é designado o **nível de significância do teste**. A escolha de  $\alpha$  depende do que se pretende estudar, pois quanto maior for  $\alpha$  mais provável será rejeitar  $H_0$  sendo esta verdadeira!!!

Uma vez recolhida a amostra, observamos o valor de alguma estatística (função da amostra aleatória) cuja distribuição de probabilidade é conhecida sob o pressuposto de  $H_0$  ser verdadeira. Tal estatística é chamada **estatística de teste**.

Nota: Qualquer estatística utilizada na construção de um intervalo de confiança para um parâmetro (variável fulcral) pode ser usada num teste de hipóteses sobre esse parâmetro.

O nível de significância do teste,  $\alpha$ , e a distribuição de probabilidade da estatística de teste vão ser utilizados para definir a chamada **região crítica** ou **região de rejeição**. Se o valor observado da estatística de teste “cair” na região crítica decidimos rejeitar  $H_0$  (e aceitar  $H_1$ ); caso contrário decidimos não rejeitar  $H_0$ .

Em resumo:

1. Formular as hipóteses do teste:  $H_0$  e  $H_1$ ;
2. Especificar um nível de significância  $\alpha$ ;
3. Escolher a estatística a usar e definir a sua distribuição de probabilidade sob o pressuposto de  $H_0$  ser verdadeira;
4. Determinar a região crítica;
5. Calcular o valor observado da estatística do teste;
6. Decidir: rejeitar  $H_0$  ou não rejeitar  $H_0$ .

#### Quadro resumo - Testes a médias e variâncias

	Condições	$H_0$	$H_1$	Estatística	Região Crítica
$\mu$	Dist. normal e $\sigma$ conhecido	$\mu = \mu_0$	$\mu \neq \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$	$Z < -z_{\alpha/2}$ ou $Z > z_{\alpha/2}$ $Z > z_\alpha$ $Z < -z_\alpha$
$\mu$	Dist. desc. e $n > 30$ com $\sigma$ conhecido (se $\sigma$ desc. usar S)	$\mu = \mu_0$	$\mu \neq \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$	$Z < -z_{\alpha/2}$ ou $Z > z_{\alpha/2}$ $Z > z_\alpha$ $Z < -z_\alpha$
$\mu$	Dist. normal e $\sigma$ desconhecido	$\mu = \mu_0$	$\mu \neq \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$	$T < -t_{\alpha/2}$ ou $T > t_{\alpha/2}$ $T > t_\alpha$ $T < -t_\alpha$
$\sigma^2$	Dist. normal	$\sigma = \sigma_0$	$\sigma \neq \sigma_0$ $\sigma > \sigma_0$ $\sigma < \sigma_0$	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi_{n-1}^2$	$\chi^2 < \chi_{1-\alpha/2; (n-1)}^2$ ou $\chi^2 > \chi_{\alpha/2; (n-1)}^2$ $\chi^2 > \chi_{\alpha; (n-1)}^2$ $\chi^2 < \chi_{1-\alpha; (n-1)}^2$

#### Exemplo

A resistência dos cabos fabricados por uma determinada empresa é uma v.a. de valor médio 816 Kg e desvio padrão 45 Kg. Adoptando-se uma nova técnica de fabrico espera aumentar-se essa resistência. Para testar tal hipótese, a empresa testou uma amostra de 50 cabos fabricados segundo o novo processo, tendo obtido uma resistência média de 839 Kg.

Pode aceitar-se a hipótese de o novo processo de fabrico aumentar a resistência dos cabos, ao nível de significância de 0.001?

## 7 Estimação por Intervalos

Como determinar um intervalo no qual se espera encontrar, com uma certa *confiança*, o valor de um parâmetro populacional  $\theta$  (ou de  $g(\theta)$ )?

### Definição

Um intervalo de confiança  $(1 - \alpha)\%$  para o parâmetro  $\theta$  (ou  $g(\theta)$ ) é um intervalo aleatório  $]L_1, L_2[$ , com  $L_1$  e  $L_2$  duas estatísticas amostrais (v.a.'s) tais que

$$P(L_1 < \theta < L_2) = 1 - \alpha.$$

$$(\theta \in ]L_1, L_2[ \text{ com probabilidade } 1 - \alpha)$$

- $1 - \alpha$  é a probabilidade do intervalo conter  $\theta$  (o grau de confiança atribuído ao intervalo);
- $\alpha \in ]0, 1[$  é a probabilidade do intervalo não conter  $\theta$  (valor preferencialmente *pequeno*).

Como calcular  $L_1$  e  $L_2$ ?

### Método da Variável Fulcral

Seja  $X$  uma v.a. cuja distribuição contém um parâmetro  $\theta$  desconhecido. Pretende-se um intervalo de confiança  $(1 - \alpha)\%$  para  $\theta$  (ou para  $g(\theta)$ ).

Considere-se uma amostra aleatória  $X_1, X_2, \dots, X_n$ , de  $X$ .

Nota: Variável Fulcral é uma função da amostra aleatória  $X_1, X_2, \dots, X_n$  e do parâmetro  $\theta$ , mas com distribuição (exacta ou aproximada) independente de  $\theta$ .

1- Escolher a Variável Fulcral (adequada)  $Z_n$

2- Determinar  $z_1, z_2 \in \mathbb{R}$ :  $P(z_1 < Z_n < z_2) = 1 - \alpha$

$$\text{Nota: } z_1 : P(Z_n < z_1) = \frac{\alpha}{2} \quad \wedge \quad z_2 : P(Z_n < z_2) = 1 - \frac{\alpha}{2}$$

3- Com  $z_1$  e  $z_2$  conhecidos, encontrar

$$L_1 \equiv L_1(X_1, X_2, \dots, X_n) \text{ e } L_2 \equiv L_2(X_1, X_2, \dots, X_n)$$

tais que

$$P(z_1 < Z_n < z_2) = 1 - \alpha \Leftrightarrow P(L_1 < \theta < L_2) = 1 - \alpha$$

$$(\theta \in ]L_1, L_2[ \text{ com probabilidade } 1 - \alpha)$$

4- Para uma amostra particular  $x_1, x_2, \dots, x_n$ , determinar estimativas para  $L_1$  e  $L_2$ , respectivamente,

$$l_1 \equiv l_1(x_1, x_2, \dots, x_n) \text{ e } l_2 \equiv l_2(x_1, x_2, \dots, x_n)$$

Obtém-se, assim, uma estimativa para o intervalo de  $\theta$ , calculado com confiança  $(1 - \alpha)\%$ ,

$$IC_\theta = ]l_1, l_2[$$

**Nota:** O intervalo  $]L_1, L_2[$  é aleatório, com probabilidade  $1 - \alpha$  de conter  $\theta$ ; isto é, se forem recolhidas amostras em número bastante elevado, espera-se que  $(1 - \alpha)\%$  dos intervalos estimados com base nessas amostras contenha  $\theta$ ; o intervalo  $]l_1, l_2[$  é apenas uma estimativa, calculado com confiança  $(1 - \alpha)\%$ , mas ao qual não se pode atribuir uma “probabilidade” de conter  $\theta$ .

## EXEMPLOS

**1-** A resistência das cordas produzidas por determinada fábrica tem distribuição normal. A fábrica testou uma amostra de 51 cordas tendo obtido uma resistência média de 300 *Kg* e desvio padrão 24 *Kg*.

- (a) Determine um intervalo de confiança 95% para a resistência média deste tipo de cordas.
- (b) Determine um intervalo de confiança 90% para o desvio padrão.

**2-** O conteúdo médio de nicotina de uma amostra de 10 cigarros de uma certa marca é de 1 *mg* (miligramas). O laboratório sabe, pela longa experiência neste tipo de análise, que o conteúdo de nicotina é uma v.a. com distribuição normal de desvio padrão 0.15 *mg*.

- (a) Determine um intervalo de confiança 90% para o conteúdo médio de nicotina.
- (b) Com a mesma confiança, quantos cigarros deveriam ser analisados de modo a que o erro máximo cometido na estimação não ultrapasse 0.01?

**3-** Realizou-se um inquérito telefónico para estimar a proporção da população de um país, favorável a determinada reforma fiscal.

No inquérito realizado a 50 pessoas, escolhidas aleatoriamente, 84% manifestou-se favorável à reforma fiscal.

Determine um intervalo de confiança 95% para a proporção da população favorável à reforma fiscal.