

ENGENHARIA ELETROTÉCNICA E DE COMPUTADORES

Probabilidade e Estatística

Apontamentos sobre PE

(Alguns tópicos 🐼)

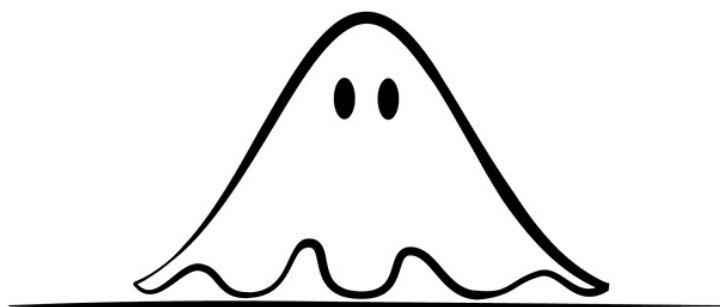


Imagem: Distribuição Paranormal

Autores:

João Gonçalves : 99995

↔ jrazevedogoncalves@tecnico.ulisboa.pt

Teresa Nogueira : 100029

↔ maria.teresa.ramos.nogueira@tecnico.ulisboa.pt

Fevereiro 2023

Índice

1. Noções Básicas de Probabilidade	1
1.1 Conceitos	1
1.1.1 Experiência Aleatória, Espaço de Resultados e Eventos	1
1.1.2 Definições de Probabilidade	2
1.1.2.1 Modelo probabilístico de Kolmogorov	3
1.2 Probabilidade condicionada	5
1.2.1 Leis das Probabilidades Compostas e da Probabilidade Total.	5
1.2.2 Teorema de Bayes	6
1.2.3 Acontecimentos Independentes	7
1.3 Notas sobre análise combinatória	8
2. Variáveis Aleatórias e Distribuições	9
2.1 Função de (densidade) probabilidade	9
2.1 Função de distribuição	10
2.2 Valor esperado, variância, moda e quantis	11
2.2.1 Valor esperado	11
2.2.2 Moda	12
2.2.2 Mediana	12
2.2.2 Quantil de probabilidade	13
2.2.3 Variância	13
2.2.4 Desvio Padrão	13
2.2.5 Coeficiente de variação	13
2.1 Distribuições discretas conhecidas	14
2.2.3 Distribuições discretas	14
2.2.3 Distribuições contínuas	18
3. Pares Aleatórios e Complementos	21
3.1 Pares Aleatórios discreto	22
3.2 Pares Aleatórios contínuos	25
3.3 Convariância e Correlação	27
3.3 Combinações lineares de variáveis aleatórias	29
3.3.1 Casos especiais de soma de v.a. independentes	30
3.3.2 Teorema do Limite Central	31
4. Amostragem e Estimação Pontual	32
4.1 Conceitos base sobre estatística	32
4.2 Método de Máxima Verosimilhança	34
5. Estimação por Intervalos	37
5.1 IC para o valor esperado e variância conhecida	38
5.2 IC para o valor esperado e variância desconhecida	39
5.3 IC para a variância e valor esperado desconhecido	40
5.4 IC's para uma probabilidade de sucesso e outros parâmetros de população não normais uniparamétricas	41
5.4.1 Probabilidade de sucesso num ensaio de Bernoulli	41
5.4.2 Valor médio de uma população Poisson	41
5.4.3 Rate de uma população exponencial	41
6. Testes de Hipóteses	42
6.1 Cálculo do valor-p	44

7. Introdução à Regressão Linear	45
7.1 Modelo da Regressão Linear Simples	45
7.1.1 Estimação de β_0 e β_1 — Método dos Mínimos Quadrados	45
7.1.2 Estimação de β_0 e β_1 — Método da Máxima Verossimilhança	46
7.1.3 Reta de Regressão	46
7.2 Intervalos de Confiança e Testes de Hipótese para β_0 , β_1 e $E(Y x = x_0)$. . .	47
7.3 Coeficiente de Determinação	47
Referências	48

1. Noções Básicas de Probabilidade

↳ Conceitos

→ Experiência Aleatória, Espaço de Resultados e Eventos

Experiência Aleatória (e.a.)

Uma experiência diz-se aleatória se

- conhecermos *a priori* todos os resultados possíveis da mesma,
- não for possível predizer o seu resultado exato antes da sua realização,
- for passível de se repetir, mesmo que hipoteticamente, nas mesmas condições ou em condições muito semelhantes.

Nota: ao invés da primeira propriedade na definição, pode referir-se que “*os resultados obtidos ao cabo de uma longa repetição da experiência patenteiam impressionante regularidade estatística quando tomadas em conjunto*”[1].

Espaço de Resultados

Conjunto de todos os resultados possíveis de uma experiência aleatória; comumente denominado por Ω .¹ Adianta-se que Ω é:

- **Discreto**, caso Ω seja (um conjunto com cardinal, $\#\Omega$) finito ou infinito numerável.
- **Contínuo**, se Ω for finito não numerável.

Evento ou Acontecimento

Subconjunto do espaço de resultados Ω . Possuem 3 classificações:

- **Elementar**, caso seja um subconjunto singular de Ω
- **Certo**, se for idêntico ao espaço de resultados Ω .
- **Impossível**, quando pertence ao subconjunto nulo.

Nota 1: Qualquer conjunto é conjunto de si próprio. Consequentemente Ω também é um evento.

Nota 2: Quando um qualquer evento A está contido (incluído) num qualquer evento B , a realização de A implica B , mas o inverso não se verifica:

realização de $A \implies$ realização de B

realização de $A \not\Leftarrow$ realização de B

Eventos Dijuntos

Quaisqueres dois eventos A e B dizem-se **dijuntos** (mutuamente exclusivos ou incompatíveis) sse a sua realização simultânea for impossível, $A \cap B = \emptyset$

¹Outras designações comuns são: *conjunto fundamental, espaço fundamental, espaço-amostra*. [1]

Operações sobre eventos

Seja Ω o espaço de resultados de uma experiência aleatória, e A e B . Então, é possível efetuar operações sobre A e B :

Operação	Notação	Descrição
Interseção	$A \cap B$	Realização simultânea de A e de B
Reunião	$A \cup B$	Realização de A ou de B , i.e., de pelo menos um dos eventos
Diferença	$B \setminus A$	Realização de B sem que se realize A (B exceto A)
Complementação	\bar{A}	Não realização de A

→ Definições de Probabilidade

Probabilidade clássica de Laplace

Considerando uma experiência aleatória com espaço de resultados Ω constituído por n eventos elementares ($\#\Omega = n$) distintos, em número finito e igualmente prováveis. Considere-se ainda que a realização do evento A passa pela ocorrência de m dos n eventos estipulados. A probabilidade de ocorrência de A é dada por:

$$P(A) = \frac{\text{número de casos favoráveis à ocorrência de } A}{\text{número de casos possíveis}} = \frac{m}{n}$$

→ Esta definição só é válida quando:

- $\#\Omega < +\infty$ (número de eventos elementares finitos)
- Ω é constituído por eventos elementares igualmente equiprováveis.

Intepretação frequencista da probabilidade

A probabilidade de um evento pode ser aproximada pela *frequência relativa* deste evento ao fim de uma grande número N de realizações da experiência aleatória (*vide* método de Monte Carlo).

$$P(A) \simeq f_N(A)$$

Def. Frequência Relativa:

$$f_N(A) = \frac{s_N(A)}{N}, \quad \text{onde } s_N \text{ é a frequência absoluta do evento.}$$

- $0 \leq f_N(A) \leq 1$
- $f_N(\Omega) = 1$
- $f_N(A)$ estabiliza (em torno de um valor fixo) à medida que N aumenta.

Nota: A estabilidade referida acima não é de natureza trivial—não é possível admitir que $f_N(A)$ estabiliza em torno de um valor, já que a repetição da e.a. é de cariz finito.²

²“Não surpreende que sintamos a tentação de (...) escrever $P(A) = \lim_{N \rightarrow +\infty} f_N(A)$. Contudo, esta igualdade trata-se [apenas] de uma *idealização matemática*.”

Limitações da interpretação frequentista da probabilidade: Só é viável se for realizada mais que uma vez em condições idênticas. Não computa casos hipotéticos.

Intepretação subjetiva de probabilidade

Um indivíduo pode atribuir a um evento um número real no intervalo $[0, 1]$, nomeado por *probabilidade subjetiva* (do acontecimento), de acordo com o grau de credibilidade que lhe associa. Isto é, coerência \iff verificação de um conjunto de *axiomas*.

Nota: esta definição sugere uma não uniformidade de processos na atribuição de *probabilidades subjetivas*. É necessário recorrer a uma definição mais *geral e rigorosa* para a noção de *probabilidade*.

★ Modelo probabilístico de Kolmogorov (1933)

Ora, o enunciado de tal definição requer que se defina a σ -álgebra sobre Ω e espaço mensurável.

σ -álgebra sobre Ω

\mathcal{A} diz-se uma σ -álgebra sobre Ω caso se trate de:

- uma coleção não vazia de subconjuntos de Ω , coleção esta não fechada para
- uniões numeráveis, interseções numeráveis e complementação.

Conjunto mínimo de propriedades que garante que \mathcal{A} é uma σ -álgebra sobre Ω :

1. $\Omega \in \mathcal{A}$;
2. $A \in \mathcal{A} \implies \bar{A} \in \mathcal{A}$, para qualquer $A \in \mathcal{A}$;
3. $\bigcup_{i=1}^{+\infty} A_i \in \mathcal{A}$, para qualquer coleção numerável $\{A_1, A_2, \dots\}$ de eventos de \mathcal{A} .

Nota: A σ -álgebra (ou *tribo de acontecimentos*) \mathcal{A} pode ser entendida como uma coleção de eventos de Ω *estável* (fechada) *para as operações de conjuntos*, na medida em que nesta pertencem todos os eventos que entendamos pertinentes. É possível definir o *evento* como qualquer conjunto A pertencente a \mathcal{A} , a σ -álgebra sobre Ω . [1]

σ -álgebra de Borel

A σ -álgebra de Borel definida para \mathbb{R} é denotada por $\mathcal{B}(\mathbb{R})$ e *gerada* pela classe de intervalos semifechados $]a, b] : -\infty < a < b < +\infty$.

Os elementos de $\mathcal{B}(\mathbb{R})$ são designados por *conjuntos de Borel* ou borelianos.

Nota: Qualquer conjunto *razoável* de \mathbb{R} — tais como conjuntos singulares ou numeráveis, intervalos fechados, abertos ou semifechados, semirretas, etc. — pertencem a $\mathcal{B}(\mathbb{R})$.

Espaço Mensurável

O par (Ω, \mathcal{A}) , constituído pelo espaço de resultados Ω dotado da σ -álgebra \mathcal{A} , é designado espaço mensurável (ou probabilizável).

Aos conjuntos pertencentes a \mathcal{A} damos o nome de eventos mensuráveis (ou probabilizáveis).

A probabilidade é uma função cujos objetos são eventos. Mas deve ser uma função σ -aditiva.

Função de Probabilidade

Seja (Ω, \mathcal{A}) um espaço mensurável. Então, uma função $P(\cdot)$ definida sobre \mathcal{A} diz-se uma função de probabilidade (no sentido de Kolmogorov), caso satisfaça os seguintes *axiomas*:

(A1) $P(A) \geq 0, \forall A \in \mathcal{A}$.

(A2) $P(\Omega) = 1$.

(A3) Seja $\{A_1, A_2, \dots\}$ uma coleção numerável de eventos disjuntos de \mathcal{A} (isto é, $A_i \cap A_j = \emptyset$). Então,

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

Espaço de Probabilidade

O terno (Ω, \mathcal{A}, P) é designado por espaço de probabilidade.

Consequências elementares dos axiomas: Os axiomas estabelecem regras para o cálculo de probabilidades,

- (i) $P(\emptyset) = 0$;
- (ii) $P(B \setminus A) = P(B) - P(A \cap B)$;
- (iii) $P(\bar{A}) = 1 - P(A)$;
- (iv) $A \in B \implies P(A) \leq P(B)$;
- (v) $P(A) \leq 1$.

Probabilidade da reunião de n eventos³

Sejam A_1, \dots, A_n eventos quaisquer. Então,

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{i=1}^n \sum_{j=i+1}^n P(A_i \cap A_j) + \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n P(A_i \cap A_j \cap A_k) \\ &\quad - \dots + (-1)^{n+1} P\left(\bigcap_{i=1}^n A_i\right) \end{aligned}$$

Para a reunião de dois eventos temos:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Para a reunião de três eventos temos:

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C) \end{aligned}$$

[§]**Nota:** Um evento pode ainda ser classificado como

1. quase-certo — se $P(A) = 1$ e, no entanto, $A \neq \Omega$;
2. quase-impossível — caso $P(A) = 0$ mas $A \neq \emptyset$.

³O resultado é designado por *fórmula de Poincaré* ou *princípio da inclusão-exclusão* (na versão probabilística).[1]

↳ Probabilidade condicionada

Probabilidade Condicionada

Sejam (Ω, \mathcal{A}, P) um espaço de probabilidade e B um evento de \mathcal{A} tal que $P(B) > 0$. A probabilidade de o evento A se realizar, sabendo que B se realizou é definida por:

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

onde $P(A|B)$ é designado pela probabilidade de A condicionada por B .

Nota: Esta probabilidade representa uma reavaliação da probabilidade A face ao facto de B ter ocorrido. Lê-se *probabilidade de A dado B* , *probabilidade de A sabendo B* , *probabilidade de A condicional à ocorrência de B* .

Probabilidade Condicionada: Considerando $P(B) > 0$ e $P_B(A) = P(A|B)$, $P_B(A)$ é uma função de probabilidade (no sentido de Kolmogorov), se satisfaz os axiomas supramencionados que constam na definição de função de probabilidade:

$$(A1') \quad P_B(A) = P(A|B) \geq 0, \forall A \in \mathcal{A}$$

$$(A2') \quad P_B(\Omega) = P(\Omega|B) = 1$$

(A3') Seja $\{A_1, A_2, \dots\}$ uma coleção numerável de eventos disjuntos de \mathcal{A} ($A_i \cap A_j = \emptyset$). Então,

$$P_B\left(\bigcup_{i=1}^{+\infty} A_i\right) = P\left(\bigcup_{i=1}^{+\infty} A_i \mid B\right) = \sum_{i=1}^{+\infty} P(A_i|B)$$

Onde P_B goza das consequências elementares da axiomática.

→ Leis das Probabilidades Compostas e da Probabilidade Total.

Lei das Probabilidades Compostas

Considere-se uma coleção de n eventos $\{A_i\}_{i=1,\dots,n}$ tal que $P(A_1 \cap A_2 \cap \dots \cap A_{n-1} \cap A_n)$. Então,

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

Partição do Espaço de Resultados Ω

A coleção de n eventos $\mathcal{P}_\Omega = \{A_i\}_{i=1,\dots,n}$ diz-se uma partição de Ω sse for constituído por eventos disjuntos, cuja reunião coincide com Ω :

- $A_i \cap A_j = \emptyset, \forall i \neq j$
- $\bigcup_{i=1}^n A_i = \Omega$

Consequentemente, caso $\mathcal{P}_\Omega = \{A_i\}_{i=1,\dots,n}$ seja uma partição de Ω :

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

Exemplo: Partições de Ω

- **E.a.:** Registo do número de acidentes na A1 durante uma ano
- **Espaço de Resultados:** $\Omega = \{0, 1, 2, \dots\}$
- **Duas partições de Ω :**

$$\mathcal{P}_\Omega = \{\{0, 2, 4, 6, \dots\}, \{1, 3, 5, \dots\}\}$$

Partição constituída pelos eventos *par* e *ímpar*

$$\mathcal{P}_\Omega = \{\{i\}\}_{i=0,1,2,\dots}$$

Partição constituída pelos eventos elementares de Ω

Lei da Probabilidade Total

Sejam:

- B um evento
- \mathcal{P}_Ω uma partição finita de Ω constituída por eventos tais que $P(A_i) > 0$ com $i = 1, \dots, n$

Então,

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

Como $\mathcal{P}_\Omega = \{A_i\}_{i=1,\dots,n}$ é uma partição do espaço de resultados Ω , é possível admitir:

$$\bigcup_{i=1}^n A_i = \Omega$$

Subsequentemente, invocando a propriedade modular das operações sobre eventos (e posteriormente a propriedade distributiva):

$$B = B \cap \Omega = B \cap \bigcup_{i=1}^n A_i = \bigcup_{i=1}^n (B \cap A_i)$$

Reconhecendo que os eventos $(B \cap A_1), \dots, (B \cap A_n)$ são disjuntos dois a dois, isto é, $(B \cap A_i) \cap (B \cap A_j) = B \cap (A_i \cap A_j) = \emptyset$ e relembrando o axioma (A3) já previamente abordado na função de probabilidade, tem-se então:

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \boxed{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

Teorema de Bayes

Sejam:

- B um evento
- \mathcal{P}_Ω uma partição de Ω tal que $P(A_i) > 0, i = 1, \dots, n$

Então,

$$P(A_i|B) \stackrel{4}{=} \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i) \times P(B|A_i)}{\sum_{j=1}^n P(B|A_j) \times P(A_j)}$$

⁴Intuitivamente, sabendo que B ocorreu, permite-nos imaginar B como o "novo" Ω . Desta forma, a probabilidade é dada pela parte de A em B, escalada para que $P(B|B) = 1$.

→ Acontecimentos Independentes

Acontecimentos Independentes

A e B dizem-se independentes sse

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

Onde é usual escrever $A \perp B$

Nota: “A independência entre eventos traduz a ausência de interação probabilística entre os mesmos”[1]. Por outro lado, a independência pode ser averiguada para quaisquer eventos (nomeadamente com probabilidade nula) ao invés da probabilidade condicionada, que é definida apenas quando $P(B) > 0$, para $P(A|B)$.

Consequências da Independência entre Eventos

1. Para qualquer evento A :
 - $A \perp \emptyset$ e $A \perp \Omega$
2. Sejam A e B independentes:
 - $P(A|B) = P(A), P(B) > 0$
 - $P(B|A) = P(B), P(A) > 0$

Ou seja, como já referido, a realização de B não influencia a concretização de A e vice-versa.

3. Sejam A e B independentes:
 - $\bar{A} \perp B, A \perp \bar{B}$ ou $\bar{A} \perp \bar{B}$

Acontecimentos Independentes Vs. Eventos Dijuntos

Sejam A e B dois eventos dijitos com probabilidades positivas:

- $A \cap B = \emptyset$
- $P(A) > 0$ e $P(B) > 0$

Então A e B não são acontecimentos independentes.

n Eventos Mutuamente Independentes

Os n eventos A_1, \dots, A_n são mutuamente independentes sse forem independentes dois a dois, três a três, ..., n a n , ou seja:

$$P\left(\bigcap_{i=1}^m A_{k_i}\right) = \prod_{i=1}^m P(A_{k_i})$$

Independência Condicional

Os eventos A e B dizem-se condicionalmente independentes em relação a C sse

$$P[(A \cap B)|C] = P(A|C) \times P(B|C)$$

Listamos abaixo alguns resultados respetivos a técnicas de contagem, cruciais no cálculo de probabilidades ao lidarmos com espaços de resultados finitos.

- Número de formas distintas de preenchimento de caixa com n compartimentos, dispõe de n elementos e não havendo a possibilidade de repetição no preenchimento. *Sequências de n elementos distintos...*

- Número de formas distintas de preenchimento de caixa com x compartimentos, dispondo de n elementos e havendo a possibilidade de repetição no preenchimento.
- Sequências de x elementos...*

- Número de formas distintas de preenchimento de caixa com x compartimentos, dispondo de n elementos e não havendo possibilidade de repetição no preenchimento.
- Sequências de x elementos distintos...*

- Número de conjuntos de cardinal x (logo com elementos distintos) que podem ser formados com n elementos. *Conjuntos de x elementos distintos...*

- Número de formas distintas de preenchimento de caixa com n compartimentos, dispondo de n elementos de k tipos distintos, onde n_i representa o número de elementos de tipo $i = 1, 2, \dots, k$, e $\sum_{i=1}^k n_i = n$.
- Sequências de n elementos de k tipos...*

- $$(a + b)^n = \sum_{x=0}^n \binom{n}{x} a^x b^{n-x}$$

[illegible]

Fig. 1: Triângulo de Pascal

⁵Ou arranjos *com* repetições.

⁶Ou arranjos *sem* repetições.

2. Variáveis Aleatórias e Distribuições

Variável Aleatória (v.a.)

Seja (Ω, \mathcal{A}) um espaço mensurável. A função real X diz-se uma variável aleatória (v.a.) se transformar o espaço de resultados Ω em \mathbb{R} e a imagem inversa de qualquer conjunto de Borel de \mathbb{R} for um evento σ -álgebra \mathcal{A} definida sobre Ω , i.e., $X : \Omega \mapsto \mathbb{R}$

$$X^{-1} = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}, \forall B \in \mathcal{B}(\mathbb{R})$$

Tipo 1: v.a. discreta — tomam valores em conjuntos numeráveis de \mathbb{R}

X diz se uma v.a. discreta caso exista um conjunto finito ou infinito numerável $D (D \subseteq \mathbb{R})$ tal que:

$$P(X \in D) = 1$$

Tipo 2: v.a. contínua — tomam valores em intervalos de \mathbb{R}

Sejam X uma v.a. e $F_X = P(X \leq x)$, $x \in \mathbb{R}$ a sua função de distribuição. X diz-se uma v.a. contínua, caso:

- $F_X(x)$ seja uma função contínua

e exista uma função real de variável real $f_X(x)$ (função densidade probabilidade) que verifique:

- $f_X(x) \geq 0 \forall x \in \mathbb{R}$
- $F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt, \forall x \in \mathbb{R}$

Nota: As últimas duas condições garantem a existência de uma função absolutamente contínua.

↳ Função de (densidade) probabilidade

Tipo 1: A v.a. discreta X com contradomínio \mathbb{R}_X toma valores x_i com probabilidades muito bem definidas. A sua caracterização somente se torna relevante se a par da indentificação do seu contradomínio adiantarmos os valores de p_i para qualquer $x_i \in \mathbb{R}_X$:

Função de probabilidade

A coleção de números $\{p_i\}$, que satisfazem

- $p_i = P(X = x_i) \geq 0$
- $\sum_i p_i = 1$

é designada de **função de probabilidade** da v.a. discreta X . Representaremos a f.p. de X por:

$$P(X = x_i) = \begin{cases} p_i, & \text{se } x = x_i \in \mathbb{R}_X \\ 0, & \text{outros valores de } x \end{cases}$$

Tipo 2: para além das propriedades já anteriormente mencionados, a **função de densidade probabilidade** satisfaz:

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1 \quad \int_b^a f_X(x)dx = P(a \leq x \leq b)$$

Por sua vez, seja X uma v.a. contínua, então:

- $P(X = x) = 0, \forall x \in \mathbb{R}$, já que o evento $X = x$ é quase impossível.
- $P(a \leq x \leq b) = P(a < x \leq b) = P(a \leq x < b) = P(a < x < b)$

Nota: ao contrário da f.p. de uma v.a. discreta, a f.p. de uma v.a. contínua não é necessariamente limitada superiormente por 1.

↳ Função de distribuição

A função de distribuição representa a probabilidade de uma dada v.a. X tomar valores não superiores a um número real arbitrário x :

Função de distribuição

A função de distribuição de uma v.a. X é dada por

$$F_X(x) = P(X \leq x) = P(x \in]-\infty, x]), x \in \mathbb{R}$$

Tipo 1: A f.d. da v.a. discreta X escreve-se "à custa" da f.p. de X :

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \sum_{\{x_i \in \mathbb{R}_X : x_i \leq x\}} P(X = x_i) \\ &= \sum_{\{x_i \leq x\}} P(X = x_i), x_i \in \mathbb{R}_X \end{aligned}$$

Propriedades da função de distribuição uma v.a. discreta

Seja X uma v.a. discreta com contradomínio \mathbb{R}_X . Então $F_X(x)$ é uma função:

- em escada, com número de pontos de descontinuidade iguais aos valores de \mathbb{R}_X .
- contínua à direita.
- monótona não decrescente de x .
- $0 \leq F_X(x) \leq 1$
- $F_X(-\infty) \rightarrow 0$ e $F_X(+\infty) \rightarrow 1$
- $P(X < x) = F_X(x^{-1}) = \lim_{h \rightarrow 0} F_X(x - h), h > 0$
- $P(a < X \leq b) = F_X(b) - F_X(a)$ e $P(a \leq X < b) = F_X(b^{-}) - F_X(a^{-})$
- $P(a \leq X \leq b) = F_X(b) - F_X(a^{-})$ e $P(a < X < b) = F_X(b^{-}) - F_X(a)$
- $P(X = x) = F_X(x) - F_X(x^{-})$

Tipo 2: Tal como no caso discreto, a f.d. da v.a. contínua X é dependente da função de probabilidade de X :

Seja X uma v.a. contínua. Então, a sua f.d é dada por

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t)dt, x \in \mathbb{R}$$

Nota: Caso F_X seja uma função absolutamente contínua e f_X uma função contínua no ponto x é possível obter a f.d.p derivando a f.d.:

$$f_X(x) = \frac{dF_X(x)}{dx}$$

Propriedades da função de distribuição uma v.a. contínua

Seja $F_X(x)$ a f.d. da v.a. contínua X . Então $F_X(x)$ é uma função:

- contínua, à direita e à esquerda $F_X(x^+) = F_X(x^-)$, $\forall x \in \mathbb{R}_X$.
- monótona não decrescente de x .
- $0 \leq F_X(x) \leq 1$
- $F_X(-\infty) \rightarrow 0$ e $F_X(+\infty) \rightarrow 1$
- $P(X \leq x) = P(X < x) = F_X(x)$
- $P(X > x) = 1 - F_X(x)$

Para a e b :

$$\begin{aligned} P(a \leq x \leq b) &= P(a < x \leq b) = P(a \leq x < b) = P(a < x < b) \\ &= \int_a^b f_X(x)dx \\ &= F_X(b) - F_X(a) \end{aligned}$$

↳ Valor esperado, variância, moda e quantis

A descrição completa do comportamento probabilístico está dependente de um conjunto de **parâmetros (indicador ou medida sumária)**:

- **Localização central** — valor esperado, moda, mediana.
- **Ordem** — quantil de probabilidade
- **Dispersão** — variância, desvio padrão e coeficiente de variação

→ Valor esperado

Tipo 1: Se $E[X] = \sum_{x \in \mathbb{R}_X} x \cdot P(X = x) < +\infty$ (se a série for absolutamente convergente) o valor esperado de uma variável aleatória discreta X com contradomínio \mathbb{R}_X existe e é dada por:

$$E[X] = \sum_{x \in \mathbb{R}_X} x \cdot P(X = x)$$

Nota: O valor esperado de uma v.a. discreta X não pertence necessariamente ao conjunto de valores possíveis de X . Desta forma, é comum verificar $E(X) \notin \mathbb{R}_X$.

Tipo 2: A definição do valor esperado no caso contínuo é análogo ao discreto. Onde tínhamos \sum e $P(X = x)$ passamos a ter \int e $f_X(X)$. O valor esperado da v.a. contínua é igual a:

$$E[X] = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx$$

Desde que o integral seja absolutamente convergente ($\int_{-\infty}^{+\infty} |x| \cdot f_X(x) dx < +\infty$).

Propriedades do valor esperado

O valor esperado satisfaz as seguintes propriedades:

1. $E[b] = b$, $b \in \mathbb{R}$.
2. $E[aX + b] = aE[X] + b$, $b, a \in \mathbb{R}$.
3. Seja $Y = \Omega(X)$ uma v.a., função da v.a. discreta X . Então, caso $E(Y)$ exista:

$$E(Y) = E(\Omega(X)) = \sum_x \Omega(X) \cdot P(X = x)$$

De forma análoga, se considerarmos o caso contínuo:

$$E(Y) = E(\Omega(X)) = \int_{-\infty}^{+\infty} \Omega(X) \cdot P(X = x)$$

De um modo geral, $E(\Omega(X)) \neq \Omega(E[X])$

→ Moda

A moda corresponde ao ponto máximo da função de probabilidade (no caso discreto) ou função de densidade probabilidade (no caso contínuo):

$$\text{tipo 1: } mo : P(mo) = \max_x f_X(x)$$

$$\text{tipo 2: } mo : P(X = mo) = \max_x P(X = x)$$

→ Mediana

A mediana de uma v.a. X é todo e qualquer ponto:

$$me : me(X) : \frac{1}{2} \leq F_X(me) \leq \frac{1}{2} + P(X = me)$$

Como no caso contínuo $P(X = me) = 0$ a fórmula reduz-se a:

$$me : me(X) : F_X(me) = \frac{1}{2}$$

→ Quantil de probabilidade

Quantil de probabilidade de p de v.a. discreta

O quantil de probabilidade p ($0 < p < 1$) da v.a. X é representado por \mathcal{X}_p e satisfaz

$$\mathcal{X}_p : p \leq F(\mathcal{X}_p) \leq p + P(X = \mathcal{X}_p)$$

No caso contínuo:

$$\mathcal{X}_p : F(\mathcal{X}_p) = p$$

→ Variância

A variância define o grau de dispersão dos valores de uma dada v.a. em torno do seu centro da gravidade:

Variância

A variância é dada por:

$$\begin{aligned} V(X) &= E\{[X - E(X)]^2\} = E\{X^2 - 2XE(X) + E(X)^2\} \\ &= E(X)^2 + E(X^2) + E[-2XE(X)] \\ &= E(X)^2 + E(X^2) + -2E(X)E(X) \\ &= E(X^2) - E(X)^2 \end{aligned}$$

e satisfaz as seguintes propriedades:

1. $E(b) = 0, b \in \mathbb{R}$.
2. $E(aX + b) = a^2V(X), b, a \in \mathbb{R}$.
3. $V(X) \geq 0$, qualquer que seja a v.a. X

→ Desvio Padrão

É costume recorrer a outra medida de dispersão absoluta, já que a variância não é expressa nas mesmas unidades que a variável aleatória. O **desvio padrão** trata-se da raiz quadrada positiva da variância de X :

$$DP(X) = +\sqrt{V(X)}$$

Independente de X ser discreta ou contínua.

→ Coeficiente de variação

O **coeficiente de variação** é uma medida de dispersão relativa, adimensional, que permite comparar a dispersão das distribuições com localizações (ou ordens de grandeza) distintas. É dada por:

$$CV(X) = \frac{DP(X)}{E(X)}$$

Independente de X ser discreta ou contínua e desde que $E[X] \neq 0$.

↳ Distribuições discretas conhecidas

→ Distribuições discretas

Distribuição uniforme discreta — Esta distribuição é razoável quando a v.a. discreta toma n valores distintos, todos com a mesma probabilidade. Sem perda de generalidade, considera-se que esta v.a. toma n valores distintos, x_1, x_2, \dots, x_n em que $x_1 < x_2 < \dots < x_n$.

Distribuição Uniforme Discreta

A v.a. discreta X diz-se ter distribuição uniforme discreta no conjunto $\{x_1, x_2, \dots, x_n\}$, se a sua f.p. coincidir com a que conta na tabela abaixo:

Notação	$X \sim \text{uniforme discreta}(\{x_1, x_2, \dots, x_n\})$
Parâmetro	$\{x_1, x_2, \dots, x_n\}$ ($x_i \in \mathbb{R}, i = 1, \dots, n$)
Contradomínio	$\{x_1, x_2, \dots, x_n\}$
F.p	$P(X = x) = \begin{cases} \frac{1}{n}, & x = x_i \in \mathbb{R}_X \\ 0, & \text{c.c.} \end{cases}$
Valor Esperado	$E(X) = \frac{1}{n} \sum_{i=1}^n x_i$
Variância	$V(X) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2$

As duas subsequentes distribuições estão dependentes da definição de **prova de Bernoulli** — Uma experiência aleatória diz-se uma prova de Bernoulli se possuir apenas dois resultados possíveis:

- Um sucesso, que ocorre com probabilidade p ($0 \leq p \leq 1$)
- Um insucesso, que ocorre com probabilidade $(1 - p)$

Distribuição de Bernoulli

Porventura, a v.a. discreta

- X = número de sucessos numa prova de Bernoulli

diz-se com **Distribuição de Bernoulli** com parâmetro p e possui a f.p. na seguinte tabela:

Notação	$X \sim \text{Bernoulli}(p)$
Parâmetro	$p = P(\text{sucesso})$ ($p \in [0, 1]$)
Contradomínio	$\{0, 1\}$
F.p	$P(X = x) = \begin{cases} p^x(1 - p)^{1-x}, & x = 0, 1 \\ 0, & \text{c.c.} \end{cases}$
Valor Esperado	$E(X) = p$
Variância	$V(X) = p(1 - p)$

Distribuição Binomial — É pertinente saber qual a distribuição do número de sucessos num número fixo n de provas de Bernoulli realizadas de forma independente e com probabilidade de sucesso comum p .

Distribuição binomial

A v.a. discreta

- X = número de sucessos num conjunto de n provas de Bernoulli independentes com probabilidade de sucesso comum e igual a p

diz-se com distribuição binomial de parâmetros (n, p) e possui a f.p. que consta na seguinte tabela:

Notação	$X \sim \text{binomial}(n, p)$
Parâmetro	n = número de provas de Bernoulli ($n \in \mathbb{N}_0$) $p = P(\text{sucesso})$ ($p \in [0, 1]$)
Contradomínio	$\{0, 1, \dots, n\}$
F.p	$P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1 \\ 0, & \text{c.c.} \end{cases}$
Valor Esperado	$E(X) = np$
Variância	$V(X) = np(1-p)$

Propriedade — Seja Y o número de insucessos em n provas de Bernoulli ($Y = n - X$)

$$\begin{aligned}
 Y &= n - X \sim \text{binomial}(n, 1 - p) \\
 F_Y(y) &= P(n - X \leq y) \\
 &= P(X \geq n - y) \\
 &= 1 - P(X < n - y) \\
 &= 1 - P(X \leq n - y - 1) \\
 &= 1 - F_X(n - y - 1)
 \end{aligned}$$

Distribuição geométrica — Caso estejamos interessados em contabilizar o número total de provas de Bernoulli realizadas até ao registo do primeiro sucesso, passamos a lidar com uma v.a. com distribuição geométrica.

Distribuição geométrica

A v.a. discreta

- X = número de provas de Bernoulli (independentes com probabilidade de sucesso comum e igual a p) até à ocorrência do primeiro sucesso.

diz-se com distribuição geométrica de parâmetro p e possui a f.p. que consta na seguinte tabela:

Notação	$X \sim \text{geométrica}(p)$
Parâmetro	$p = P(\text{sucesso})$ ($p \in [0, 1]$)
Contradomínio	$\{1, 2, \dots, n\}$
F.p	$P(X = x) = \begin{cases} p(1-p)^{x-1}, & x = 1, 2, \dots \\ 0, & \text{c.c.} \end{cases}$
Valor Esperado	$E(X) = \frac{1}{p}$
Variância	$V(X) = \frac{1-p}{p^2}$

Falta de memória da distribuição geométrica — Seja $X \sim \text{geométrica}(p)$. Então,

$$P(X > k + x | X > k) = P(X > x), \forall k, x \in \mathbb{N}$$

De modo equivalente, a v.a. $(X - k | X > k)$ que representa o número de provas adicionais, sabendo que não se registou qualquer sucesso nas primeiras k provas, também possui distribuição geométrica com parâmetro p :

$$(X - k | X > k) \sim \text{geométrica}(p), k \in \mathbb{N}$$

Esta propriedade traduz um **recomeço probabilístico**. A informação acerca da não ocorrência de sucessos nas primeira k provas de Bernoulli não é relevante para os cálculos subsequentes.

Distribuição Poisson — Frequentemente usada na contagem de ocorrências de certo tipo de eventos em períodos fixos de tempo, áreas, volumes ... Eventos tais como chegadas, acidentes, falhas de equipamento, testemunhos verdadeiros em tribunal, etc.

Distribuição poisson

A v.a. X diz-se com distribuição de Poisson e parâmetro λ caso possua f.p. na tabela seguinte:

Notação	$X \sim \text{Poisson}(\lambda)$
Parâmetro	$\lambda \ (\lambda \in \mathbb{R}^+)$
Contradomínio	\mathbb{N}_0
F.p	$P(X = x) = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!}, & x = 0, 1, 2, \dots \\ 0, & \text{c.c.} \end{cases}$
Valor Esperado	$E(X) = \lambda$
Variância	$V(X) = \lambda$

Nota: Se porventura procurarmos a distribuição da expansão ou contração da unidade para a qual é realizada a contagem de ocorrências, basta multiplicar o parâmetro λ pelo coeficiente da transformação em causa.

A aproximação de Poisson da distribuição binomial — A pertinência desta aproximação prende-se com o facto de nem sempre ser possível calcular a probabilidade de certos eventos recorrendo às tabelas disponíveis:

A f.p. da v.a. $X \sim \text{binomial}(np)$ pode ser aproximada satisfatoriamente pela f.p./f.d. da v.a. aproximativa

$$\tilde{X} \sim \text{Poisson}(np), \quad \text{quando } n > 20 \text{ e } p < 0.1$$

já que

$$\lim_{\substack{n \rightarrow +\infty \\ p \rightarrow 0 \\ np = \lambda \text{ fixo}}} \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda} \frac{\lambda^x}{x!}$$

Adianta-se também que:

$$\sum_{x=0}^{+\infty} \left| \binom{n}{x} p^x (1-p)^{n-x} - e^{-\lambda} \frac{\lambda^x}{x!} \right| \leq 2p, \quad \text{para } 0 \leq p \leq 1$$

→ Distribuições contínuas

Distribuição uniforme contínua — Esta distribuição é o análogo da distribuição uniforme discreta. Não surpreende que se trate de uma distribuição adequada a descrever o comportamento probabilístico de uma v.a. cujos valores possíveis se crê terem todos o mesmo peso.

Distribuição Uniforme Contínua

A v.a. contínua X diz-se ter distribuição uniforme contínua no intervalo $[a, b]$, se a sua f.p. coincidir com a que conta na tabela abaixo:

Notação	$X \sim \text{uniforme}(a, b)$
Parâmetro	a extremo inferior do intervalo; $a \in \mathbb{R}$ b extremo superior do intervalo; $b \in \mathbb{R}$
Contradomínio	$[a, b]$
F.d.p	$f_X(x) = \begin{cases} \frac{1}{b-a}, & xa \leq x \leq b \\ 0, & \text{c.c.} \end{cases}$
Valor Esperado	$E(X) = \frac{a+b}{2}$
variância	$V(X) = \frac{(a+b)^2}{12}$

Propriedade — Considere-se que $X \sim \text{uniforme}(a, b)$. Então, os intervalos — com a mesma amplitude e contidos em $[a, b]$ — são equiprováveis.

Distribuição Normal — Esta distribuição surge associada à modelação de observações relativas a medições de temperaturas, velocidades, erros, volumes de ruído, etc. Surge também como distribuição aproximada, nomeadamente de algumas distribuições discretas, ou ainda, de médias aritméticas ou somas de v.a.

Distribuição Normal

A v.a. contínua X diz-se ter distribuição normal de parâmetros μ e σ^2 caso possua a seguinte f.d.p:

Notação	$X \sim \text{normal}(\mu, \sigma^2)$
Parâmetro	μ ($\mu \in \mathbb{R}$) σ^2 ($\sigma^2 \in \mathbb{R}^+$)
Contradomínio	\mathbb{R}
F.d.p	$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$
Valor Esperado	$E(X) = \mu$
Variância	$V(X) = \sigma^2$

F.d.p da v.a. normal — A f.d.p da v.a. normal é simétrica em torno de μ e tem forma de sino. É devido a estes dois factos que a mediana é igual à moda e ao valor esperado:

$$me(X) = mo(X) = E(X) = \mu$$

O valor de σ^2 , como seria de esperar, determina o achatamento da f.d.p. desta v.a.:

- σ^2 pequeno, f.d.p muito alongada
- σ^2 grande, f.d.p muito achatada

F.d. da v.a. normal — A f.d. da v.a. normal é dada por:

$$\int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt, \quad -\infty < x < +\infty$$

Não possui expressão fechada, pelo que só pode ser obtida numericamente.

Distribuição Normal Padrão — Seja $X \sim \text{normal}(\mu, \sigma^2)$. Então, recorrendo ao processo de estandarização:

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - \mu}{\sigma} \sim \text{normal}(0, 1)$$

$$F_Z(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \Phi(z)$$

Z possui **distribuição normal padrão**. É à custa de Z que se obtém a f.d. da v.a. $X \sim \text{normal}(\mu, \sigma^2)$, para quaisquer valores admissíveis de μ e σ^2

$$F_X(x) = P(X \leq x) = P\left(Z = \frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

Nota: Graças à simetria da f.p.d. da normal padrão em torno da origem, podemos concluir que

$$\Phi(-z) = 1 - \Phi(z), \quad -\infty < x < +\infty$$

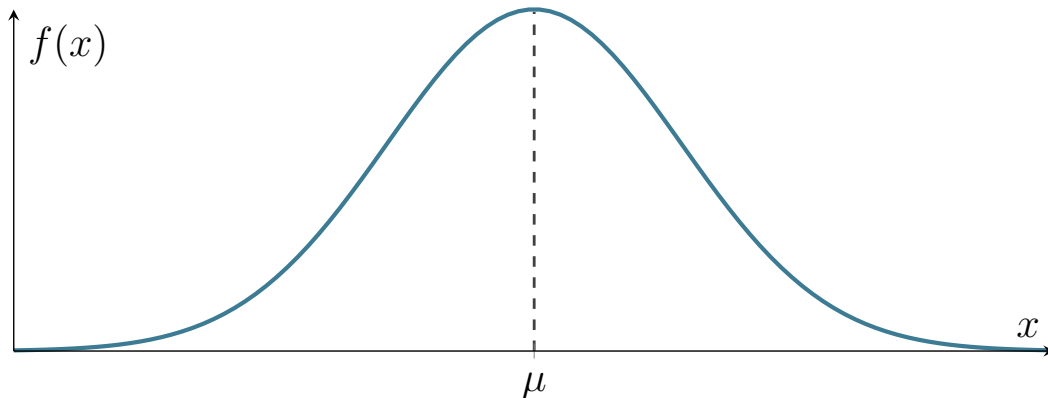


Fig. 2: Função Densidade de Probabilidade (f.d.p) da Distribuição Normal — $\text{normal}(\mu, \sigma^2)$

Distribuição exponencial — É a distribuição mais utilizada na caracterização da duração de equipamento, naquilo que usualmente é designado por *testes de vida*.

Distribuição exponencial

A v.a. contínua X diz-se ter distribuição exponencial de parâmetro μ caso possua a seguinte f.d.p:

Notação	$X \sim \text{exponencial}(\lambda)$
Parâmetro	$\lambda \ (\lambda \in \mathbb{R}^+)$
Contradomínio	\mathbb{R}_0^+
F.d.p	$f_X(x) = \lambda e^{-\lambda x}, \ x \geq 0$
Valor Esperado	$E(X) = \frac{1}{\lambda}$
Variância	$V(X) = \frac{1}{\lambda^2}$

Falta de memória da distribuição exponencial — Seja $X \sim \text{exponencial}(\lambda)$. Então,

$$\begin{aligned} P(X \leq t + x | X > t) &= P(X \leq x), \forall t, x \in \mathbb{R}_0^+ \\ P(X > t + x | X > t) &= P(X > x), \forall t, x \in \mathbb{R}_0^+ \\ (X - t | X > t) &\sim \text{exponencial}(\lambda), t \in \mathbb{R}_0^+ \end{aligned}$$

Ao considerarmos que X representa a vida de um item e que X tem distribuição exponencial(λ), a vida residual no instante t , $(X - t | X > t)$ possuirá a mesma distribuição e o mesmo parâmetro de X .

Relação entre as distribuições exponencial e de Poisson — Sejam:

- N_t o número de ocorrências de um evento no intervalo $]0, t]$, $t > 0$
- X_i o tempo entre ocorrências consecutivas $(i - 1)$ e i do evento

Caso a coleção de v.a. $\{N_t, t > 0\}$ seja um processo de Poisson de taxa λ , é sabido que $N_t \sim \text{Poisson}(\lambda t)$ e podemos concluir que

$$X_i \stackrel{i.i.d}{\sim} \text{exponencial}(\lambda), \ i \in \mathbb{N}$$

O de $\stackrel{i.i.d}{\sim}$ lê-se *independentes e identicamente distribuídos a*

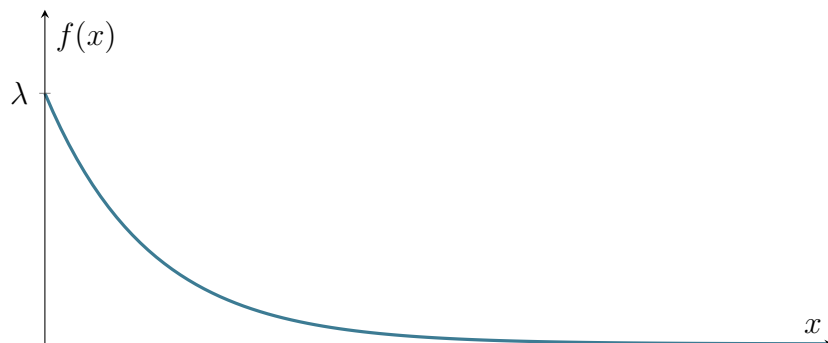


Fig. 3: Função Densidade de Probabilidade (f.d.p) da Distribuição Exponencial — exponencial(λ)

3. Pares Aleatórios e Complementos

Pares Aleatórios — A definição de par aleatório (ou v.a. bidimensional) é análoga à de v.a. unidimensional: lidamos com uma função — com características especiais — que transforma eventos em pares ordenados.

Par Aleatório

A função (X, Y) diz-se um par aleatório caso transforme o espaço Ω em \mathbb{R}^2 e a imagem inversa — segundo (X, Y) — de qualquer região do tipo $] - \infty, x] \times] - \infty, y]$ pertença à σ -álgebra \mathcal{A} definida sobre Ω

$$(X, Y) : (X, Y)^{-1} = (] - \infty, x] \times] - \infty, y]) \in \mathcal{A}, \forall (x, y) \in \mathbb{R}^2$$

De notar que ao considerarmos $(X, Y)(\omega)$, a imagem inversa integrante na condição de mensurabilidade é a seguinte:

$$(X, Y)^{-1} = (] - \infty, x] \times] - \infty, y]) = \{\omega \in \Omega : X(\omega) \leq x, Y(\omega) \leq y\}$$

Eis dois exemplos de pares aleatórios:

- O número de sinais emitidos (X) por um aparelho e o número desses sinais que foram recebidos (Y) por um aparelho recetor — par de v.a. contínuas.
- A distância (X) de um ponto de impacto ao centro de um alvo e o ângulo (Y) com o eixo das abcissas — par de v.a. contínuas.

Função de distribuição conjunta de um par aleatório — A f.d. conjunta de (X, Y) é dada por:

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

Propriedades da função de distribuição conjunta de um par aleatório — A f.d. conjunta satisfaz as seguintes propriedades:

1. função contínua à direita com respeito aos argumentos x e y .
2. função monótona não decrescente em qualquer das variáveis x e y .
3. $0 \leq F_{X,Y}(x, y) \leq 1$
4. $F_{X,Y}(-\infty, -\infty) = \lim_{x,y \rightarrow -\infty} F_{X,Y}(x, y) = 0$
 $F_{X,Y}(-\infty, y) = \lim_{x \rightarrow -\infty} F_{X,Y}(x, y) = 0$
 $F_{X,Y}(x, -\infty) = \lim_{y \rightarrow -\infty} F_{X,Y}(x, y) = 0$
 $F_{X,Y}(+\infty, +\infty) = \lim_{x,y \rightarrow \infty} F_{X,Y}(x, y) = 1$
5. $F_{X,Y}(x, +\infty) = \lim_{y \rightarrow +\infty} F_{X,Y}(x, y) = F_X(x)$
 $F_{X,Y}(-\infty, y) = \lim_{x \rightarrow -\infty} F_{X,Y}(x, y) = F_Y(y)$

↳ Pares Aleatórios discreto

Tipo 1: Par Aleatório discreto — toma valores em conjuntos finitos ou infinitos numeráveis de \mathbb{R}^2

(X, Y) diz-se um par aleatório discreto caso tome valores exclusivamente num conjunto finito ou infinito numerável $\mathbb{R}_{X,Y} = \{(x_i, y_j)\}_{i=1,2,\dots, j=1,2,\dots}$:

$$P[(X, Y) \in \mathbb{R}_{X,Y}] = 1$$

Função de probabilidade conjunta — Seja (X, Y) um par aleatório discreto com contradomínio $\mathbb{R}_{X,Y}$. Então, a f.p. conjunta de (X, Y) é definida por:

$$\begin{aligned} P(X = x, Y = y) &= P(\{\omega \in \Omega : X(\omega) = x, Y(\omega) = y\}) \\ &= \begin{cases} P(X = x_i, Y = y_j) & \text{se } (x, y) = (x_i, y_j) \in \mathbb{R}_{X,Y} \\ 0, & \text{c.c.} \end{cases} \end{aligned}$$

E raramente é representada graficamente. Quando o contradomínio é finito e não muito numeroso, é convenientemente disposta numa tabela do tipo abaixo:

X / Y	y_1	y_2	y_3
x_1	$P(X = x_1, Y = y_1)$	$P(X = x_1, Y = y_2)$	$P(X = x_1, Y = y_3)$
x_2	$P(X = x_2, Y = y_1)$	$P(X = x_2, Y = y_2)$	$P(X = x_2, Y = y_3)$
x_3	$P(X = x_3, Y = y_1)$	$P(X = x_3, Y = y_2)$	$P(X = x_3, Y = y_3)$

Tab. 1: Possível função de probabilidade conjunta de X e Y .

Função de distribuição conjunta

A f.d. conjunta do par aleatório discreto (X, Y) pode ser obtida à custa da f.p. conjunta:

$$\begin{aligned} F_{X,Y}(x, y) &= P(X \leq x, Y \leq y) \\ &= \sum_{x_i \leq x} \sum_{y_j \leq y} P(X = x_i, Y = y_j), \quad (x, y) \in \mathbb{R}^2 \end{aligned}$$

Função de probabilidade marginais de X e Y

A f.p. marginal do par aleatório discreto (X, Y) pode ser obtida à custa da f.p. conjunta:

$$P(X = x) = \sum_y P(X = x, Y = y), \quad x \in \mathbb{R}$$

A f.p. marginal de Y obtém-se de modo análogo:

$$P(Y = y) = \sum_x P(X = x, Y = y), \quad x \in \mathbb{R}$$

Função de distribuição marginais de X e Y

A f.d. marginais do par aleatório discreto (X, Y) podem ser obtidas à custa da f.p. marginal:

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} \left[\sum_y P(X = x_i, Y = y) \right] = F_{X,Y}(x, +\infty)$$

$$F_Y(y) = P(Y \leq y) = \sum_{y_i \leq y} \left[\sum_x P(X = x, Y = y_i) \right] = F_{X,Y}(+\infty, y)$$

Função de probabilidade condicional — De modo a verificar a forma como o registo de uma observação de uma das v.a. do par aleatório influencia o comportamento probabilístico de outra v.a. recorreremos à função de probabilidade condicional.

Função de probabilidade condicional de X e Y

A f.p. condicional a $Y = y$ — caso $P(Y = y) > 0$:

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}, \quad x \in \mathbb{R}$$

A f.p. condicional a $X = x$ — caso $P(X = x) > 0$:

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}, \quad x \in \mathbb{R}$$

De modo semelhante, é conveniente definir as funções de distribuição condicionais:

Função de distribuição condicional de X e Y

A f.d. condicional a $Y = y$ — caso $P(Y = y) > 0$:

$$F_{X|Y=y}(x) = P(X \leq x|Y = y) = \sum_{x_i \leq x} \frac{P(X = x_i, Y = y)}{P(Y = y)}, \quad x \in \mathbb{R}$$

A f.d. condicional a $X = x$ — caso $P(X = x) > 0$:

$$F_{Y|X=x}(y) = P(Y \leq y|X = x) = \sum_{y_i \leq y} \frac{P(X = x, Y = y_i)}{P(X = x)}, \quad x \in \mathbb{R}$$

Bem como os respetivos parâmetros de caracterização:

- Valor esperado de $X|Y = y$

$$E(X = x|Y = y) = \sum_x x P(X = x|Y = y) \quad \text{se } P(Y = y) > 0$$

- Valor esperado de $Y|X = x$

$$E(Y = y|X = x) = \sum_y y P(Y = y|X = x) \quad \text{se } P(X = x) > 0$$

Admitindo que $E(X^2)$ e $E(Y)^2$ existem:

- Variância de $X|Y = y$

$$\begin{aligned} V(X|Y = y) &= E(X^2|Y = y) - E(X|Y = y)^2 \\ &= \sum_x x^2 P(X = x|Y = y) - \left[\sum_x x P(X = x|Y = y) \right]^2 \end{aligned}$$

desde que $P(Y = y) > 0$

- Variância de $Y|X = x$

$$\begin{aligned} V(Y|X = x) &= E(Y^2|X = x) - E(Y|X = x)^2 \\ &= \sum_y y^2 P(X = x|Y = y) - \left[\sum_y y P(X = x|Y = y) \right]^2 \end{aligned}$$

desde que $P(X = x) > 0$

(In)dependência — De um modo geral, as v.a. que constituem um par aleatório influenciam-se mutuamente, são **dependentes**. Caso tal não aconteça dizem-se **independentes**.

Independência

A variáveis dizem-se independentes — escrevendo-se $X \perp\!\!\!\perp Y$ — se

$$P(X = x, Y = y) = P(X = x) \times P(Y = y), \quad \forall (x, y) \in \mathbb{R}_{X,Y}$$

Ou seja, caso seja possível escrever a f.p. conjunta do par aleatório discreto (X, Y) à custa do produto das f.p. marginais de X e Y , para qualquer par ordenado $(x, y) \in \mathbb{X}, \mathbb{Y}$

Podemos enunciar a noção de independência entre duas v.a. da seguinte forma:

$$P(X \in B_X, Y \in B_Y) = P(X \in B_X) \times P(Y \in B_Y)$$

para quaisquer conjuntos de Borel B_X e B_Y de \mathbb{R} Podemos definir independência entre as v.a. X e Y (com qualquer carácter) do seguinte modo. As v.a. X e Y dizem-se independentes sse

$$F_{X,Y}(x, y) = F_X(x) \times F_Y(y), \quad \forall (x, y) \in \mathbb{R}^2$$

Caso $X \perp\!\!\!\perp Y$, sse for possível escrever a f.d. conjunta do par aleatório (X, Y) à custa do produto das f.d. marginais de X e Y . Além disso, caso $X \perp\!\!\!\perp Y$ e $P(X = x|Y = y)$ esteja definida, temos:

- $P(X = x|Y = y) = P(X = x)$
- $E(X|Y = y) = E(X)$
- $V(X|Y = y) = V(X)$

↳ Pares Aleatórios contínuos

Tipo 2: Par Aleatório contínuo — Tal como no caso unidimensional, a definição de par aleatório contínuo é sustentada na definição de f.d.p. conjunta.

Par aleatório contínuo e função de densidade de probabilidade conjunta

O par aleatório (X, Y) diz-se contínuo se tomar valores num conjunto infinito não numerável $\mathbb{R}_{X,Y} \subset \mathbb{R}^2$, e existir uma função denominada de f.d.p. conjunta, $f_{X,Y}(x, y)$, satisfazendo:

- $f_{X,Y}(x, y) \geq 0 \quad \forall (x, y) \in \mathbb{R}^2$
- $F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) dv du \quad (x, y) \in \mathbb{R}^2$

Propriedades da função de densidade probabilidade conjunta — A f.d.p. conjunta $f_{X,Y}(x, y)$ satisfaz:

- $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(u, v) dv du = 1$
- $f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}$, caso $f_{X,Y}$ seja contínua no ponto (x, y)
- $P[(X, Y) \in A] = \iint_A f_{X,Y}(u, v) dv du$, para qualquer evento A de \mathbb{R}^2

Função de distribuição conjunta

A f.d. conjunta do par aleatório contínuo (X, Y) define-se à custa da respetiva f.d.p. conjunta:

$$\begin{aligned} F_{X,Y} &= P(X \leq x, Y \leq y) \\ &= \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) dv du \quad (x, y) \in \mathbb{R}^2 \end{aligned}$$

Onde $F_{X,Y}(x, y)$ corresponde, obviamente, ao volume sob a superfície da f.d.p. conjunta (X, Y) na região $] - \infty, x] \times] - \infty, y]$

A função de distribuição conjunta é tipicamente representada mediante uma função por ramos.

Função de densidade de probabilidade marginais de X e Y

As f.d.p. marginais de X e Y definem-se à custa da respetiva f.d.p. conjunta:

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy \\ f_Y(y) &= \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx \end{aligned}$$

Funções de distribuição marginais de X e Y — de forma análoga ao caso discreto, as f.d. marginais de X e de Y calculam-se à custa das respetivas f.d.p. marginais, ou, então por recurso à f.d.p. conjunta ou à f.d. conjunta do par aleatório contínuo (X, Y) .

Função de distribuição marginais de X e Y

$$\begin{aligned} F_X(x) &= P(X \leq x) = \int_{-\infty}^x f_X(x)dx = \int_{-\infty}^x \left[\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy \right] dx \\ &= F_{X,Y}(x, +\infty), \quad x \in \mathbb{R} \\ F_Y(y) &= P(Y \leq y) = \int_{-\infty}^y f_Y(y)dy = \int_{-\infty}^y \left[\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx \right] dy \\ &= F_{X,Y}(+\infty, y), \quad y \in \mathbb{R} \end{aligned}$$

Funções de densidade de probabilidade condicional — Ao lidarmos com um par aleatório contínuo (X, Y) é absurdo definir a probabilidade condicionada $P(X = x|Y = y)$, já que $P(X = x) = P(Y = y) = 0$, no entanto, faz sentido definir:

F.d.p, F.d, valores esperados e variância condicionais

Admitindo $f_Y(y) > 0$

- F.d.p. de $X|Y = y$

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad x \in \mathbb{R}$$

- F.d. de $X|Y = y$

$$F_{X|Y=y} = \int_{-\infty}^x f_{X|Y=y}(x)dx, \quad x \in \mathbb{R}$$

Assumindo que $E(X)$ existe:

- Valor esperado de $X|Y = y$

$$E(X|Y = y) = \int_{-\infty}^{+\infty} x f_{X|Y=y}(x)dx$$

Admitindo também que $E(X^2)$ existe:

- Variância de $X|Y = y$

$$\begin{aligned} V(X|Y = y) &= E(X^2|Y = y) - E(X|Y = y)^2 \\ &= \int_{-\infty}^{+\infty} x^2 f_{X|Y=y}(x)dx - \left[\int_{-\infty}^{+\infty} x f_{X|Y=y}(x)dx \right]^2 \end{aligned}$$

Independência

As v.a. contínuas X e Y dizem-se independentes sse

$$f_{X,Y}(x,y) = f_X(x) \times f_Y(y), \quad \forall (x,y) \in \mathbb{R}^2$$

Ou seja, se formos capazes de escrever a f.d.p. conjunta do par aleatório contínuo (X, Y) à custa do produto das f.d.p. marginais de X e Y . Caso $X \perp\!\!\!\perp Y$ e $F_{X,Y}(x,y)$ esteja definida verificamos as mesmas propriedades que no caso discreto, com exceção:

$$F_{X|Y=y}(x,y) = f_X(x)$$

↳ Covariância e Correlação

É crucial obter medidas que avaliam a associação (linear) entre duas v.a.:

- de forma absoluta, calculando a covariância entre X e Y
- de forma relativa, determinando a correlação entre X e Y

Covariância

Sejam X e Y duas v.a. para as quais existem momentos de 2ª ordem $E(X^2)$ e $E(Y^2)$. Então, a covariância entre X e Y representa-se usualmente por $cov(X, Y)$ e é definida por

$$\begin{aligned} cov(X, Y) &= E\{[X - E(X)] \times [Y - E(Y)]\} \\ &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ &= E(XY) - 2 \cdot E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

Se X e Y forem v.a. independentes então a covariância entre X e Y é nula, mas o inverso não é necessariamente verdadeiro. Mais, se $cov(X, Y) \neq 0$, podemos concluir imediatamente que X e Y são v.a. dependentes.

Propriedades da Covariância

Sejam X, Y, Z, X_1, \dots, X_n e Y_1, \dots, Y_n v.a. com segundos momentos finitos. Então:

1. $X \perp\!\!\!\perp Y \implies cov(X, Y) = 0$
2. $cov(X, Y) = 0 \not\Rightarrow X \perp\!\!\!\perp Y$
3. $cov(X, Y) \neq 0 \implies X \not\perp\!\!\!\perp Y$
4. $cov(X, Y) = cov(Y, X)$
5. $cov(X, X) = V(X)$
6. $cov(aX + b, Y) = acov(X, Y)$
7. $cov(X + Z, Y) = cov(X, Y) + cov(Z, Y)$
8. $cov(\sum_{i=1}^n X_i, \sum_{j=1}^n Y_j) = \sum_{i=1}^n \sum_{j=1}^n cov(X_i, Y_j)$

A covariância é uma medida absoluta de associação, como tal possui algumas desvantagens:

- Não é adimensional
- O seu valor pode ser alterado de modo arbitrário, ao efetuar-se, por exemplo uma mudança de escala (vide propriedade 6 na tabela de propriedades da covariância).

Há pois necessidade de quantificar, em termos relativos, a associação entre X e Y à custa de uma medida adimensional e passível de interpretação:

Correleação

Sejam X e Y duas v.a. para asquais existem momentos de 2ª ordem $E(X^2)$ e $E(Y^2)$. Então, a correlação entre X e Y representa-se usualmente por $\text{corr}(X, Y)$ e é definida por

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{V(X) \cdot V(Y)}}$$

Caso $\text{corr}(X, Y) \neq 0$ (repetivamente $\text{corr}(X, Y) = 0$) as variáveis aleatórias dizem-se correlacionadas (respetivamente não correlacionadas).

Propriedades da Convariância

Sejam X e Y v.a. com segundos momentos finitos. Então:

1. $X \perp\!\!\!\perp Y \implies \text{corr}(X, Y) = 0$
2. $\text{corr}(X, Y) = 0 \not\implies X \perp\!\!\!\perp Y$
3. $\text{corr}(X, Y) \neq 0 \implies X \not\perp\!\!\!\perp Y$
4. $\text{corr}(X, Y) = \text{corr}(Y, X)$

Importa ainda adiantar:

1. $\text{corr}(X, X) = 1$
2. $\text{corr}(aX + b, Y) = \begin{cases} -\text{corr}(X, Y), & a < 0 \\ \text{corr}(X, Y) & a > 0 \end{cases}$
3. $-1 \leq \text{corr}(X, Y) \leq 1$ para qualquer par de v.a.
4. $|\text{corr}(X, Y)| = 1 \rightarrow Y = aX + b$ (quase em toda a parte com $\text{corr}(X, Y) = -1$ se $a < 0$ e $\text{corr}(X, Y) = 1$ se $a > 0$)

Intepertação do sinal da correlação — As propriedades 7 e 8 sugerem:

- A correlação quantifica a associação linear entre X e Y

Assim, se o valor absoluto de $\text{corr}(X, Y)$ estiver próximo de 1, podemos afirmar que a associação entre X e Y é praticamente linear.

Por seu lado o sinal da correlação deve ser interpretado do seguinte modo:

- Caso $\text{corr}(X, Y) > 0$ (resp. $\text{corr}(X, Y) < 0$) pode afirmar-se que as v.a. de X e Y tenderão a variar no mesmo sentido (resp. em sentidos opostos) relativamente aos respetivos valores esperados.

Nota: São de evitar, no entanto, interpretações abusivas desta medida de associação relativa, pois a existência de correlação entre duas v.a. não implica necessariamente uma relação direta de causa e efeito (por exemplo, o número de carros que passam numa estrada e o número de gatos que roubam peixes do mercado por mês)

↳ Combinações lineares de variáveis aleatórias

Combinação linear de v.a.

Sejam X_1, \dots, X_n v.a. e c_1, \dots, c_n constantes reais. Então, a v.a.

$$Y = \sum_{i=1}^n c_i X_i$$

diz-se uma combinação linear das v.a. X_1, \dots, X_n .

A obtenção da f.(d.)p. de uma combinação linear nem sempre é tarefa fácil. Podemos, no entanto, adiantar o valor esperado de qualquer combinação linear de v.a. e obter a variância de combinações lineares de forma expedita em situações particulares.

Valor esperado de combinações lineares de v.a.

Sejam X_1, \dots, X_n v.a. com valor esperado finito e c_1, \dots, c_n constantes reais. Então

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

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

Variância de combinações lineares de v.a.

Sejam X_1, \dots, X_n v.a. com segundos momentos finitos e c_1, \dots, c_n constantes reais. Então

$$V(c_1 X_1 + c_2 X_2) = c_1^2 V(X_1) + c_2^2 V(X_2) + 2c_1 c_2 \text{cov}(X_1, X_2)$$

$$V(X_1 + X_2) = V(X_1) + V(X_2) + 2\text{cov}(X_1, X_2)$$

$$V(X_1 - X_2) = V(X_1) + V(X_2) - 2\text{cov}(X_1, X_2)$$

$$V\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i^2 V(X_i) + 2 \sum_{i=1}^n \sum_{j=i+1}^n c_i c_j \text{cov}(X_i, X_j)$$

Ao lidarmos com v.a. não correlacionadas duas a duas (isto é, v.a. tais que $\text{cov}(X_i, X_j) = 0, \forall i \neq j$) ou com v.a. independentes duas a duas (ou seja $X_i \perp\!\!\!\perp X_j, \forall i \neq j$), temos:

$$V\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i^2 V(X_i)$$

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i)$$

→ Casos especiais de soma de v.a. independentes

Soma de v.a. Bernoulli independentes

$$X_i \sim \text{Bernoulli}(p) \quad \forall i = 1, \dots, n$$

Sejam X_i e X_j variáveis independentes para $i \neq j$, nomeadamente,

$$P(X_i = x, X_j = y) = P(X_i = x)P(X_j = y), \quad \forall x, y$$

Então,

$$\sum_{i=1}^n X_i \sim \text{binomial}(n, p)$$

Nota: As v.a. dizem-se identicamente distribuídas, já que possuem todas o mesmo parâmetro e distribuição.

Soma de v.a. Binomiais independentes

$$X_i \sim \text{Binomial}(n_i, p) \quad \forall i = 1, \dots, n$$

Então,

$$\sum_{i=1}^n X_i \sim \text{binomial}\left(\sum_{i=1}^n n_i, p\right)$$

Nota: Graças à propriedade reprodutiva da distribuição binomial, ao lidar-se com v.a. binomiais independentes com probabilidade de sucesso comum.

Soma de v.a. Poisson independentes

$$X_i \sim \text{Poisson}(\lambda_i) \quad \forall i = 1, \dots, n$$

Então,

$$\sum_{i=1}^n X_i \sim \text{Poisson}\left(\sum_{i=1}^n \lambda_i\right)$$

Nota: Graças à propriedade reprodutiva da distribuição Poisson, ao lidar-se com v.a. Poisson independentes.

Combinações lineares de v.a. normais independentes

$$X_i \sim \text{normal}(\mu_i, \sigma_i^2) \quad \forall i = 1, \dots, n$$

Então,

$$a_i X_i + b_i \sim \text{normal}(a_i \mu + b_i, a_i^2 \sigma_i^2)$$

$$\sum_{i=1}^n (a_i X_i + b_i) \sim \text{normal}\left(\sum_{i=1}^n (a_i \mu + b_i), \sum_{i=1}^n (a_i^2 \sigma_i^2)\right)$$

→ Teorema do Limite Central

Nem todas as distribuições gozam da propriedade reprodutiva inerente às distribuições previamente abordadas, pelo que é crucial saber como obter os valores aproximados de probabilidades de eventos envolvendo *somas* e *médias aritméticas* de v.a. independentes e identicamente distribuídas (i.i.d).

Para tal, recorreremos ao **Teorema do Limite Central** (TLC).

Teorema do Limite Central

Considere-se que as v.a. X_1, \dots, X_n são i.i.d com valor esperado μ e variância finita positiva σ para todo o inteiro positivo n . Considere-se, ainda, a soma das n primeiras v.a., $S_n = \sum_{i=1}^n X_i$:

$$\lim_{n \rightarrow +\infty} P \left[\frac{S_n - E(S_n)}{\sqrt{V(S_n)}} \leq z \right] = \lim_{n \rightarrow +\infty} P \left(\frac{S_n - \mu}{\sqrt{n}\sigma} \leq z \right) = \Phi(z)$$

Nota: Este é um resultado assintótico que poderá ser usado para amostras de **tamanho superior a 30** e é válido para **somas e médias aritméticas** de v.a. quer discretas, quer contínuas. Possui a seguinte fórmula abreviada:

$$\frac{\sum_{i=1}^n X_i - \sum_{i=1}^n E(X_i)}{\sqrt{\sum_{i=1}^n V(X_i)}} \sim \text{normal}(0, 1)$$

e pode também ser escrito à custa de médias, já que

$$\frac{n}{n} \cdot \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n E(X_i)}{\sqrt{\sum_{i=1}^n V(X_i)}} = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \stackrel{a}{\sim} \text{normal}(0, 1)$$

(onde $\stackrel{a}{\sim}$ deve ler-se *tem distribuição aproximadamente*)

4. Amostragem e Estimação Pontual

↳ Conceitos base sobre estatística

De seguida enumeram-se de forma sucinta, um conjunto de conceitos referentes à estatística:

População e Amostra

População — conjunto de todas as observações possíveis de determinada variável de interesse X

Amostra — subconjunto de valores da população. Seleccionada de forma aleatória e representativa da população de onde foi retirada.

Inferência estatística — O estudo de uma amostra de uma população é designada por inferência estatística. são admitidos dois níveis de ignorância em relação ao conhecimento da população em estudo:

- $F_X(x)$ é **completamente desconhecida**, sabendo-se apenas se é do tipo contínuo ou discreto.
- Admite-se (pelo conhecimento dos fenómenos em causa) que $F_X(x)$ pertence a determinada família, mas com **parâmetros desconhecidos**

Amostragem aleatória — Para que as inferências sejam rigorosas é natural que o processo de recolha de informação seja fruto do acaso.

Amostra aleatória

Sejam:

- X uma v.a. de interesse
- X_1, \dots, X_n v.a. independentes e identicamente distribuídas (i.i.d) a X , i.e., $X_i \stackrel{i.i.d}{\sim} X, i = 1, \dots, n$ ($n \in \mathbb{N}$).

Então o vetor aleatório

- $\underline{X} = (X_1, \dots, X_n)$ diz-se uma amostra aleatória (a.a.) de dimensão n proveniente da população X

Amostra

A uma observação particular da a.a. $\underline{X} = (X_1, \dots, X_n)$ dá-se o nome de amostra e representa-se por $\underline{x} = (x_1, \dots, x_n)$

Nota: Convém recordarmos que a a.a. $\underline{X} = (X_1, \dots, X_n)$ é um vetor n -dimensional e que o mesmo não acontece com a amostra $\underline{x} = (x_1, \dots, x_n)$, que é um vetor de \mathbb{R}^n

Caracterização da amostra aleatória

A caracterização da a.a. é realizada sem grande dificuldade, já que é constituída por n v.a. independentes e identicamente distribuídas de X . Neste sentido temos:

- Discreto — f.p. conjunta de \underline{X}

$$\begin{aligned} P(\underline{X} = \underline{x}) &= P(X_1 = x_1, \dots, X_n = x_n) \\ &\stackrel{X_i \text{ indep}}{=} \prod_{i=1}^n P(X_i = x_i) \\ &\stackrel{X_i \simeq X}{=} \prod_{i=1}^n P(X = x_i) \end{aligned}$$

- Contínuo — f.d.p. conjunta de \underline{X}

$$\begin{aligned} P(\underline{X} = \underline{x}) &= f_{X_1, \dots, X_n}(x_1, \dots, x_n) \\ &\stackrel{X_i \text{ indep}}{=} \prod_{i=1}^n f_{X_i}(x_i) \\ &\stackrel{X_i \simeq X}{=} \prod_{i=1}^n f_X(x_i) \end{aligned}$$

Valor esperado e Variância — Neste seguimento, o cálculo do valor médio e variância para a média da a.a. é efetuado do seguinte modo (denotando que as n v.a. são i.i.d):

Valor esperado:

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} \cdot n \cdot E(X) \\ &= E(X) \end{aligned}$$

Variância:

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

Estatística — É fundamental e conveniente condensar a informação amostral (dados) em medidas sumárias como a média, o desvio padrão da amostra ou outras medidas estudadas em estatística descritiva. Estas medidas mais não são que valores particulares de v.a. definidas à custa da a.a. e denominadas estatísticas.

Estatística

Seja $\underline{X} = (X_1, \dots, X_n)$ uma a.a. de dimensão n proveniente da população X . Então T diz-se uma estatística caso se trate de uma função (Borel mensurável) da a.a., $T = T(\underline{X})$, que não envolva qualquer parâmetro desconhecido.

Alguns Exemplos de estatísticas:

- **Média amostral**

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

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

- **Variância amostral**

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

A variância de uma amostra concreta

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

é um valor observado da v.a. S^2

- **Mínimo da a.a.:** $X(1) = \min\{X_1, X_2, \dots, X_n\}$
- **Máximo da a.a.:** $X(n) = \max\{X_1, X_2, \dots, X_n\}$

↳ Método de Máxima Verosimilhança

Estimador — É indispensável adiantar valores razoáveis para parâmetros desconhecidos que integram a nossa v.a. de interesse. Para tal, iremos recorrer a estatísticas especiais de nome *estimadores*

Estimadores

A estatística $T = T(\underline{X})$ diz-se estimador do parâmetro desconhecido θ , caso a estatística $T = T(\underline{X})$ tome valores exclusivamente no espaço paramétrico Θ .

Estimativa

Ao valor observado do estimador $T = T(\underline{X})$ do parâmetro desconhecido θ , $t = T(\underline{x})$, damos o nome de estimativa de θ .

Método de Máxima Verosimilhança — Este método permite obter o valor mais plausível/verossímil de um parâmetro desconhecido de entre todos os valores possíveis para esse mesmo parâmetro desconhecido — tendo em conta a amostra $\underline{x} = (x_1, \dots, x_n)$.

O parâmetro θ é desconhecido, mas x_1, \dots, x_n são conhecidos. Então, qual é a melhor aproximação para θ ? O número que maximiza a probabilidade/densidade de obter a amostra realmente observada. O valor de θ mais compatível com os dados observados!

Por forma a descrever o método da MV é necessário definir a função de verosimilhança

Função de verosimilhança

É representada por $L(\theta|\underline{x})$, dá ideia de quão plausível é o valor θ para o parâmetro desconhecido, caso se tenha recolhido a amostra \underline{x} e define-se do seguinte modo:

- no caso discreto, é dada por

$$L(\theta|\underline{x}) = P(\underline{X} = \underline{x} | \theta) = \prod_{i=1}^n P(X = x_i | \theta), \quad \theta \in \Theta$$

- No caso contínuo é igual a

$$L(\theta|\underline{x}) = f_{\underline{X}}(\underline{x} | \theta) = \prod_{i=1}^n f_X(x_i | \theta), \quad \theta \in \Theta$$

Onde $P(\bullet|\theta)$ e $f_X(\bullet|\theta)$ representa a f.p. e a f.d.p (resp.) da v.a. de interesse X , tendo em conta que θ é o verdadeiro valor do parâmetro desconhecido.

Quer o parâmetro desconhecido, quer o valor que se lhe possa atribuir são tradicionalmente representados por θ .

$L(\theta|\underline{x}) : \Theta \rightarrow \mathbb{R}$, i.e., a função de verosimilhança tem por argumento θ , possui por domínio o espaço paramétrico Θ e toma valores em \mathbb{R} , para cada valor fixo da amostra de \underline{x} .

Função de verosimilhança

Obtida a amostra $\underline{x} = (x_1, \dots, x_n)$, a estimativa de MV do parâmetro desconhecido θ — representada doravante por $\hat{\theta}$ — corresponde ao ponto máximo da função de verosimilhança, i.e.,

$$\hat{\theta} : L(\hat{\theta}|\underline{x}) = \max_{\theta \in \Theta} L(\theta|\underline{x})$$

Função log-verosimilhança; estimativa de máxima verosimilhança

1. É usual designar $\ln L(\theta|\underline{x})$ por função *log-verosimilhança*
2. É analiticamente mais conveniente obter o ponto máximo de $\ln L(\theta|\underline{x})$ (uma soma de logaritmos), que determinar o ponto máximo de $L(\theta|\underline{x})$ (um produto). Uma vez que o logaritmo é uma função estritamente crescente em \mathbb{R}^+ , podemos definir a estimativa de MV como o ponto máximo da função log-verosimilhança.

$$\hat{\theta} : \ln L(\hat{\theta}|\underline{x}) = \max_{\theta \in \Theta} \ln L(\theta|\underline{x})$$

3. Caso o espaço paramétrico Θ seja contínuo, recorreremos ao procedimento usual de maximização — começamos por obter o ponto de estacionaridade para, de seguida, averiguarmos se tal ponto é efetivamente um ponto máximo. Ao lidarmos com um único parâmetro desconhecido, temos:

$$\hat{\theta} : \left. \frac{d \ln L(\hat{\theta} | \underline{x})}{d\theta} \right|_{\theta=\hat{\theta}} = 0 \quad (\text{ponto de estacionaridade})$$

$$\left. \frac{d^2 \ln L(\hat{\theta} | \underline{x})}{d\theta^2} \right|_{\theta=\hat{\theta}} < 0 \quad (\text{ponto de máximo})$$

4. Caso o espaço paramétrico Θ seja discreto, a estimativa de MV de θ obtém-se calculando os vários valores de $L(\hat{\theta} | \underline{x})$ para $\theta \in \Theta$, e identificando o ponto de máximo — i.e., **faz-se por pesquisa ponto por ponto**.

A estimativa de MV é, naturalmente, uma função da amostra, i.e.,

$$\hat{\theta} = g(\underline{x})$$

Além disso, não se trata de uma v.a. mas sim de uma concretização de uma v.a. com um nome particular *estimador de MV*:

Estimador de máxima verosimilhança

O estimador de MV de θ obtém-se por substituição de $\underline{x} = (x_1, \dots, x_n)$ por $\underline{X} = (X_1, \dots, X_n)$ na expressão geral da estimativa de MV de θ , $\hat{\theta} = g(\underline{x})$. Ele será doravante representado por

$$\text{EMV}(\theta) = g(\underline{X})$$

Trata-se de uma v.a. exclusivamente dependente da a.a. \underline{X} , logo, uma estatística.

Propriedade de invariância do estimador de máxima verosimilhança

Sejam:

- $\hat{\theta}$ a estimativa de MV de θ .
- $\text{EMV}(\theta)$ o estimador de MV de θ .
- $h(\theta)$ uma função bijetiva.

Então, a estimativa de MV de $h(\theta)$ é dada por

$$\widehat{h(\theta)} = h(\hat{\theta})$$

e o estimador de MV de $h(\theta)$ por

$$\text{EMV}[h(\theta)] = h[\text{EMV}(\theta)]$$

Nota: O estimador MV de θ não é único caso lidemos, por exemplo, com um modelo uniparamétrico. Supondo $\underline{X} \sim \text{uniforme}(\theta - 1/2, \theta + 1/2) : \theta \in \mathbb{R}$, a respetiva função MV pode ser escrita como:

$$L(\theta | \underline{x}) = \begin{cases} 1 & \max(x_1, \dots, x_n) - 1/2 \leq \theta \leq \min(x_1, \dots, x_n) + 1/2 \\ 0 & \text{cc} \end{cases}$$

Podemos selecionar qualquer valor no intervalo $[\max(x_1, \dots, x_n) - 1, \min(x_1, \dots, x_n)]$ como estimativa MV para θ . O estimador de MV não é unicamente especificado.

5. Estimação por Intervalos

Intervalo de Confiança — É importante adiantar um intervalo de valores razoáveis para θ que dê uma ideia da confiança que se pode depositar na estimativa pontual de θ . Este intervalo será definido oportunamente e é denominado de intervalo de confiança (**IC**). Os valores mais usuais para o grau de confiança de um **IC** são 90%, 95% e 99%.

Intervalo (aleatório) de confiança

Admitimos que $\underline{X} = (X_1, \dots, X_n)$ representa uma a.a. proveniente da população X , cuja distribuição depende do parâmetro desconhecido θ ($\theta \in \Theta$). Sejam $T_1 = T(\underline{X})$ e $T_2 = T_2(\underline{X})$ duas estatísticas tais que $T_1 < T_2$ e

$$P(T_1 \leq \theta \leq T_2) = 1 - \alpha, \quad \forall \theta \in \Theta$$

onde $\alpha \in (0, 1)$. O intervalo

$$[T_1, T_2] = [T_1(\underline{X}), T_2(\underline{X})]$$

é designado intervalo aleatório de confiança (**IAC**) a $(1 - \alpha) \cdot 100\%$ para θ . Ao dispormos de uma amostra $\underline{x} = (x_1, \dots, x_n)$, obtemos uma concretização de $[T_1(\underline{X}), T_2(\underline{X})]$, seja ela

$$[t_1, t_2] = [T_1(\underline{x}), T_2(\underline{x})]$$

que designamos por intervalo de confiança a $(1 - \alpha) \cdot 100\%$ para θ e prepresentaremos por $IC_{(1-\alpha) \cdot 100\%}(\theta)$

É necessário adiantar um método de obtenção sistemático de IC para um parâmetro aleatório θ . Para tal, será necessário introduzir a noção de **v.a. fulcral**

V.a. fulcral

Uma v.a. que depende exclusivamente da a.a. \underline{X} e do parâmetro desconhecido θ

$$Z = Z(\underline{X}, \theta)$$

diz-se uma v.a. fulcral para θ caso possua distribuição (exata ou aproximada) independente de θ , ou de qualquer outro parâmetro desconhecido que possa existir.

Método da v.a. fulcral

Antes de tudo é crucial descrever a situação com que lidamos, nomeadamente, v.a. de interesse e respetiva distribuição; o parâmetro desconhecido alvo do IC; outros parâmetros conhecidos (ou não) da distribuição. Posto isto:

• Passo 1 — Seleção da variável fulcral

A v.a. fulcral para θ é por regra uma função trivial dos estimador de MV θ .

• Passo 2 — Obtenção dos quantis de probabilidade

O par de quantis depende do grau de confiança $(1 - \alpha) \cdot 100\%$ e será representado por a_α e b_α . De um modo geral:

$$(a_\alpha, b_\alpha) : \begin{cases} P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2 \end{cases} \iff \begin{cases} a_\alpha = F_Z^{-1}(\alpha/2) \\ b_\alpha = F_Z^{-1}(1 - \alpha/2) \end{cases}$$

(Continuação)

- **Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$**

De modo a obtermos um IAC que contenha $\{\theta\}$ com probabilidade $(1 - \alpha)$ é crucial invertermos a dupla desigualdade $a_\alpha \leq Z \leq b_\alpha$ em ordem a θ

$$P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha \rightarrow P[T_1(\underline{X}), T_2(\underline{X})] = 1 - \alpha$$

onde T_1 e T_2 são dois extremos aleatórios dependentes quer da a.a. \underline{X} , quer dos quantis de probabilidade a_α e b_α

- **Passo 4 — Concretização**

Basta agora substituir as v.a. pelas respectivas observações, x_1, \dots, x_n para obter o $IC_{(1-\alpha) \cdot 100\%}(\theta)$.

$$IC_{(1-\alpha) \cdot 100\%}(\theta) = [T_1(\underline{x}), T_2(\underline{x})]$$

↳ IC para o valor esperado e variância conhecida

Supondo inicialmente que a v.a. de interesse possui **distribuição normal**, temos:

- **Situação**

$$X \sim \text{normal}(\mu, \sigma^2) \rightarrow \begin{cases} \mu - \text{DESCONHECIDO} \\ \sigma^2 - \text{CONHECIDO} \end{cases}$$

- **Passo 1 — Seleção da v.a. fulcral para μ**

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

- **Passo 2 — Obtenção dos quantis de probabilidade**

$$(a_\alpha, b_\alpha) : \begin{cases} a_\alpha = \Phi(\alpha/2) = -\Phi(1 - \alpha/2) \\ b_\alpha = \Phi(1 - \alpha/2) \end{cases}$$

- **Passo 3 — Inversão de igualdades**

$$P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha$$

$$P\left[a_\alpha \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq b_\alpha\right] = 1 - \alpha$$

$$P\left[\bar{X} - b_\alpha \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + a_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

- **Passo 4 — Concretização**

$$IC_{(1-\alpha) \cdot 100\%} = \left[\bar{x} \pm \Phi(1 - \alpha/2) \cdot \frac{\sigma}{\sqrt{n}}\right]$$

Caso a população possua **distribuição arbitrária** (e uma amostra suficientemente grande, respetivamente ≥ 30) o procedimento é idêntico ao enunciado acima, com exceção da v.a. fulcral e respetivo IC, que passam agora a ser denominados por *aproximados* (consequência direta da utilização do TLC).

- **Passo 1 — Seleção da v.a. fulcral para μ**

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \stackrel{a}{\sim} \text{normal}(0, 1)$$

- **Passo 4 — Concretização**

$$IC_{(1-\alpha) \cdot 100\%} \simeq \left[\bar{x} \pm \Phi(1 - \alpha/2) \cdot \frac{\sigma}{\sqrt{n}} \right]$$

↳ IC para o valor esperado e variância desconhecida

Supondo inicialmente que a v.a. de interesse possui **distribuição normal**, temos:

- **Situação**

$$X \sim \text{normal}(\mu, \sigma^2) \rightarrow \begin{cases} \mu - \text{DESCONHECIDO} \\ \sigma^2 - \text{DESCONHECIDA} \end{cases}$$

- **Passo 1 — Seleção da v.a. fulcral para μ**

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

- **Passo 2 — Obtenção dos quantis de probabilidade**

$$(a_\alpha, b_\alpha) : \begin{cases} a_\alpha = F_t(\alpha/2) = -F_t(1 - \alpha/2) \\ b_\alpha = F_t(1 - \alpha/2) \end{cases}$$

- **Passo 3 — Inversão de igualdades**

$$\begin{aligned} P(a_\alpha \leq Z \leq b_\alpha) &= 1 - \alpha \\ P\left[a_\alpha \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq b_\alpha\right] &= 1 - \alpha \\ P\left[\bar{X} - b_\alpha \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + a_\alpha \cdot \frac{S}{\sqrt{n}}\right] &= 1 - \alpha \end{aligned}$$

- **Passo 4 — Concretização**

$$IC_{(1-\alpha) \cdot 100\%} = \left[\bar{x} \pm F_t(1 - \alpha/2) \cdot \frac{s}{\sqrt{n}} \right]$$

Caso a população possua **distribuição arbitrária** (e uma amostra suficientemente grande, respetivamente ≥ 30) o procedimento é idêntico ao enunciado acima, com exceção da v.a. fulcral, que passa agora a ter distribuição aproximadamente normal, ao invés de *t-student* (consequência direta da utilização do TLC). $\implies Z = \frac{\bar{X} - \mu}{S/\sqrt{n}} \stackrel{a}{\sim} \text{normal}(0, 1)$

↳ IC para a variância e valor esperado desconhecido

Supondo inicialmente que a v.a. de interesse possui **distribuição normal**, (caso a distribuição da v.a. de interessa não seja dada, admitimos sempre distribuição normal) temos:

- Situação

$$X \sim \text{normal}(\mu, \sigma^2) \rightarrow \begin{cases} \mu - \text{DESCONHECIDO} \\ \sigma^2 - \text{DESCONHECIDA} \end{cases}$$

- Passo 1 — Seleção da v.a. fulcral para μ

$$Z = \frac{S^2(n-1)}{\sigma} \sim \chi_{n-1}^2 \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

- Passo 2 — Obtenção dos quantis de probabilidade

$$(a_\alpha, b_\alpha) : \begin{cases} a_\alpha = F_{\chi^2}(\alpha/2) \\ b_\alpha = F_{\chi^2}(1 - \alpha/2) \end{cases}$$

Já que a distribuição *qui-quadrado* possui assimetria positiva.

- Passo 3 — Inversão de igualdades

$$\begin{aligned} P(a_\alpha \leq Z \leq b_\alpha) &= 1 - \alpha \\ P\left[a_\alpha \leq \frac{S^2(n-1)}{\sigma} \leq b_\alpha\right] &= 1 - \alpha \\ P\left[\frac{(n-1)S^2}{b_\alpha} \leq \sigma^2 \leq \frac{(n-1)S^2}{a_\alpha}\right] &= 1 - \alpha \end{aligned}$$

- Passo 4 — Concretização

$$IC_{(1-\alpha) \cdot 100\%} = \left[\frac{(n-1)S^2}{F_{\chi^2}(1 - \alpha/2)} ; \frac{(n-1)S^2}{F_{\chi^2}(\alpha/2)} \right]$$

↳ IC's para uma probabilidade de sucesso e outros parâmetros de população não normais uniparamétricas

Importa adiantar IC aproximados para a probabilidade de sucesso de uma prova de Bernoulli, bem como outros parâmetros de populações não normais uniparamétricas. Os passos de cálculo para o intervalo de tempo em nada se alteram relativamente aos realizados nas secções anteriores. Neste sentido, é relevante evidenciar apenas as variáveis fulcrais interatuantes no cálculo do IC:

→ Probabilidade de sucesso num ensaio de Bernoulli

$$Z = \frac{\bar{X} - p}{\sqrt{\bar{X}(1 - \bar{X})/n}} \stackrel{a}{\sim} \text{normal}(0, 1)$$

Desde que o tamanho da amostra justifique o uso do teorema do limite central. Note-se ainda que $E(\bar{X}) = E(X) = p$, $V(\bar{X}) = V(X)/n = p(1 - p)/n \simeq \bar{X}(1 - \bar{X})/n$ de acordo com a distribuição binomial.

→ Valor médio de uma população Poisson

$$Z = \frac{\bar{X} - \lambda}{\sqrt{\bar{X}/n}} \stackrel{a}{\sim} \text{normal}(0, 1)$$

Novamente, Desde que o tamanho da amostra justifique o uso do teorema do limite central, onde $V(\bar{X}) = \lambda/n \simeq \bar{X}/n$.

→ Rate de uma população de exponencial

Recorrendo ao Teorema do Limite Central:

$$\begin{aligned} Z &= \frac{\bar{X} - \lambda^{-1}}{\lambda^{-1}/\sqrt{n}} \\ &= (\lambda\bar{X} - 1) \cdot \sqrt{n} \stackrel{a}{\sim} \text{normal}(0, 1) \end{aligned}$$

6. Testes de Hipóteses

O teste de hipótese visa verificar se uma dada amostra contém ou não informação que deve ser aceite (ou não) no cálculo de parâmetros desconhecidos.

Hipótese paramétrica

Trata-se de conjectura sobre um parâmetro desconhecido θ , assumindo que se conhece a distribuição da v.a. de interesse X a menos de θ

Hipótese nula/alternativa

Ao confrontarmos duas hipóteses paramétricas:

- a hipótese mais relevante é usualmente designada por hipótese nula, é representada por H_0 e correspondente a $H_0 : \theta \in \Theta_0$
- $H_1 : \theta \in \Theta_1$ é designada por hipótese alternativa

Nomeadamente, por exemplo:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

e neste caso a hipótese alternativa é denominada de **bilateral**

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu > \mu_0 \text{ ou } \mu < \mu_0$$

e neste caso a hipótese alternativa é denominada de **unilateral**

Erros de 1ª e 2ª espécie (tipo I e tipo II)

As probabilidades de ocorrência de erros da 1ª e 2ª espécie costumam ser representadas por α ou β e definem-se do seguinte modo:

$$\alpha = P(\text{Erro de 1ª espécie}) = P(\text{Rejeitar } H_0 | H_0)$$

$$\beta = P(\text{Erro de 2ª espécie}) = P(\text{Não rejeitar } H_0 | H_1)$$

Nível de significância

O teste de hipóteses deve ser delineado de modo a que se verifique

$$P(\text{Rejeitar } H_0 | H_0) \leq \alpha_0$$

Onde α_0 é designado por nível de significância (congruente com o grau de confiança).

Região de Rejeição de H_0 (para valores de estatística de teste) — É habitual escolhermos esta região de modo a que:

- W satisfaça a condição $P(\text{Rejeitar } H_0 | H_0) = P(T \in W | H_0) = \alpha_0$ (onde T é designada por estatística de teste).
- W seja um intervalo real (ou uma reunião de intervalos reais) que depende de quantis de probabilidade relacionada com α_0 e que digam respeito à distribuição exata ou aproximada da estatística de teste sob a validade de H_0 .
- O aspeto de H_0 depende da hipótese alternativa H_1 e do que significa obter valores inconsistentes com H_0 .

É relevante expor um esqueleto para o procedimento geral de testes de hipótese:

Procedimento geral de testes de hipótese

Efetuar um teste de hipóteses compreende os sete passos seguintes:

1. V.a. de interesse

Identificar a v.a. de interesse

2. Situação

Tal como no cálculo dos intervalos de confiança, é necessário adiantar a distribuição da v.a. de interesse, o parâmetro desconhecido que está a ser testado, bem como outros possíveis parâmetros desconhecidos (ou conhecidos), com vista a escolher a variável fulcral apropriada.

3. Hipóteses

Enunciar as hipóteses nula e alternativa, de forma análoga à especificada acima.

4. Nível de significância

Escolher o nível de significância (tipicamente dado, caso contrário recorrer ao *p-value*)

5. Estatística de teste

Selecionar a estatística de teste adequada (T) e identificar a sua distribuição (exata ou aproximada) sob H_0

$$T \stackrel{H_0}{\sim} \text{distribuição}$$

Onde T é definida de forma análoga à realizada nos intervalos de confiança, com recurso à variável fulcral adequada (cujo o processo de escolha em nada difere do capítulo anterior).

6. Região de Rejeição

Identificar a região de rejeição de H_0 para valores da estatística de teste:

$$\begin{cases} c = F_Z^{-1}(1 - \alpha_0/2) & \text{caso bilateral }] -\infty, -c [\cup] c, +\infty [\\ c = F_Z^{-1}(\alpha_0) & \text{caso unilateral esquerdo }] -\infty, -c [\\ c = F_Z^{-1}(1 - \alpha_0) & \text{caso unilateral direito }] c, \infty [\end{cases}$$

7. Decisão

Calcular o valor observado da estatística de teste (t) e decidir pela rejeição ou não de H_0 ao nível de significância α_0 .

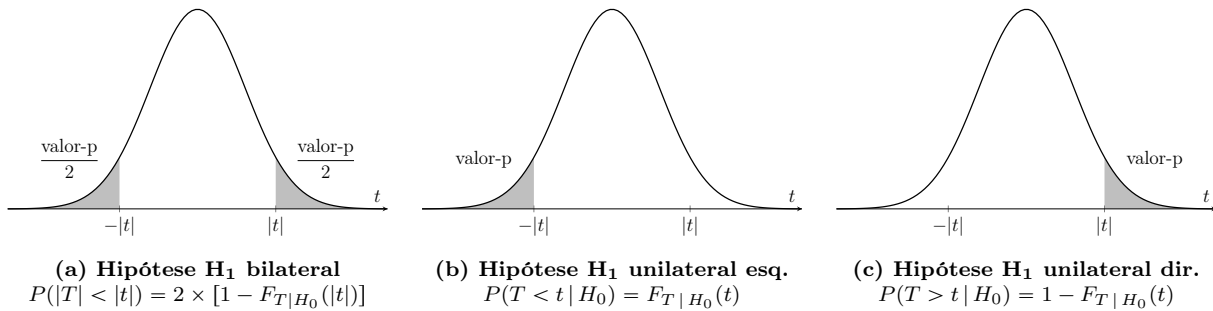
Nota: Uma qualquer hipótese que seja rejeitado para um dado nível de significância, será rejeitado para qualquer outro superior a este. De forma idêntica, qualquer hipótese que não seja rejeitado para um dado nível de significância, não será rejeitado para qualquer outro inferior a este.

↳ Cálculo do valor-p

Valor-p

O valor-p é o maior nível de significância que leva à não rejeição de H_0

O cálculo do valor-p depende do aspeto da região de H_0 (para valores da estatística de teste) e da concretização da estatística de teste, (valor que delimita a área de rejeição), tal como se evidencia abaixo:



Nota: Caso a distribuição da v.a. de interesse não seja simétrica, (e.g. qui-quadrado χ^2), e a hipótese alternativa seja bilateral, o valor-p passa a ser definido da seguinte maneira:

$$\text{valor-p} = 2 \times \min\{P(T < t | H_0), P(T > t | H_0)\}$$

Teste de ajustamento do qui-quadrado (hipótese nula simples)

Quando conjecturamos uma única distribuição para a v.a. de interesse X , a estatística a utilizar neste teste de ajustamento é

$$T = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \underset{a}{\sim}_{H_0} \chi_{k-1}^2$$

onde k , O_i e E_i representam o número de classes que constam na tabela de frequências, a frequência absoluta observável da classe i e a frequência absoluta esperada sob H_0 da classe i respetivamente. Salienta-se que:

$$O_i \sim \text{binomial}(n, p_i) \rightarrow E_i = nP(X \in \text{classe} | H_0) = np_i^0$$

A região de rejeição da hipótese nula simples H_0 é o intervalo à direita onde

$$W =]c, +\infty[\quad \text{onde} \quad c = F_{\chi_{k-1}^2}^{-1}(1 - \alpha_0)$$

pois quanto maiores as discrepâncias entre frequências absolutas observadas e esperadas sob H_0 das classes, maior é o valor observado da estatística de teste e menos consistente é a hipótese H_0 com os dados recolhidos

7. Introdução à Regressão Linear

↳ Modelo da Regressão Linear Simples

É frequente estarmos interessados em estabelecer uma relação linear entre:

- Uma variável dependente ou resposta, Y_i (aleatória)
- Uma variável independente (ou de controlo), X_i (considera-se não aleatória)

A equação deste modelo é dada por

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n$$

onde:

- Y_i = i -ésima resposta aleatória (v.a. dependente);
- x_i = i -ésima observação da variável explicativa ou regressora;
- β_0 = ordenada na origem do valor esperado da resposta (desconhecida);
- β_1 = declive do valor esperado da resposta (desconhecida);
- ε_i = erro aleatório associado à observação da i -ésima resposta.

1º. conjunto de hipóteses de trabalho do modelo de regressão linear simples:

- É costume assumir que os erros aleatórios ε_i são v.a. não correlacionadas, i.e.,

$$\text{corr}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j$$

De tal forma que $E(\varepsilon_i) = 0$ e $V(\varepsilon_i) = \sigma^2$, onde σ^2 é uma constante desconhecida. Consequentemente, para $i = 1, \dots, n$

$$\begin{aligned} E(Y_i) &= \beta_0 + \beta_1 x_i + E(\varepsilon_i) = \beta_0 + \beta_1 x_i \\ V(Y_i) &= V(\varepsilon_i) = \sigma^2 \end{aligned}$$

→ Estimação de β_0 e β_1 — Método dos Mínimos Quadrados

“A obtenção das estimativas dos mínimos quadrados de β_0 e β_1 passa pela minimização dos desvios entre o que é observado, y_i , e o que é *esperado* de acordo com o modelo de RLS, $E(Y_i) = \beta_0 + \beta_1 x_i$.”[1]

Queremos encontrar as $\hat{\beta}_0$ e $\hat{\beta}_1$ que minimizem

$$Q = Q(\beta_0, \beta_1) = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2$$

a soma dos quadrados dos desvios (verticais) entre y_i e $\beta_0 + \beta_1 x_i$.

$$(\hat{\beta}_0, \hat{\beta}_1) : \begin{cases} \frac{\partial Q}{\partial \beta_0} \Big|_{\beta_0=\hat{\beta}_0, \beta_1=\hat{\beta}_1} = 0 \\ \frac{\partial Q}{\partial \beta_1} \Big|_{\beta_0=\hat{\beta}_0, \beta_1=\hat{\beta}_1} = 0 \end{cases} \implies \begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \end{cases}$$

→ **Estimação de β_0 e β_1 — Método da Máxima Verosimilhança**

As estimativas de MV de β_0 e β_1 requerem um **2º. conjunto de hipóteses** de trabalho

- Assume-se, doravante, que

$$\varepsilon_i \stackrel{\text{i.i.d.}}{\sim} \text{normal}(0, \sigma^2), \quad i = 1, \dots, n$$

Então, para $i = 1, \dots, n$:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \stackrel{\text{indep}}{\sim} \text{normal}(\beta_0 + \beta_1 x_i, \sigma^2)$$

$$f_{Y_i|\beta_0, \beta_1}(y_i) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}[y_i - (\beta_0 + \beta_1 x_i)]^2\right\}$$

1. Função de verosimilhança

$$\begin{aligned} L(\beta_0, \beta_1 | \underline{x}, \underline{y}) &= \prod_{i=1}^n f_{Y_i|\beta_0, \beta_1}(y_i) \\ &= \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}[y_i - (\beta_0 + \beta_1 x_i)]^2\right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}[y_i - (\beta_0 + \beta_1 x_i)]^2\right\} \end{aligned}$$

2. Função de log-verosimilhança

$$\begin{aligned} \ln L(\beta_0, \beta_1 | \underline{x}, \underline{y}) &= -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2 \\ &= -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{\sigma^2} Q \end{aligned}$$

em que Q é como definido anteriormente.

3. Maximização

Note-se que $\ln L(\beta_0, \beta_1 | \underline{x}, \underline{y})$ é uma função proporcional a $(-Q)$, e a sua maximização de β_0 e β_1 é equivalente à maximização de Q que condiziu às estimativas

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{e} \quad \hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$$

→ **Reta de Regressão**

É usual estimar o valor esperado da resposta associada a um valor arbitrário x da variável explicativa. A estimativa pontual de $E(Y|x) = \beta_0 + \beta_1 x$ é igual a

$$\hat{y} = \widehat{E(Y|x)} = \hat{\beta}_0 + \hat{\beta}_1 x, \quad \text{para } x \in [x_{(1)}, x_{(n)}]$$

onde $x_{(1)} = \min_{i=1, \dots, n} x_i$, e $x_{(n)} = \max_{i=1, \dots, n} x_i$. A reta é usualmente designada por reta de regressão, reta ajustada ou *estimated regression line*.

↳ Intervalos de Confiança e Testes de Hipótese: $\beta_0, \beta_1, E(Y|x = x_0)$

$$1) T = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}}}$$

v.a. fulcral se:

- i) $IC_{(1-\alpha)}[\beta_0]$
- ii) Teste de hipótese para β_0

$$2) T = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}}}$$

v.a. fulcral se:

- i) $IC_{(1-\alpha)}[\beta_1]$
- ii) Teste de hipótese para β_1

Nota: Testar a significância do modelo de regressão linear, $H_0 : \beta_1 = 0$, verifica se existe relação linear significativa entre y e x (se o x não impacta o y)

$$3) T = \frac{(\hat{\beta}_0 + \hat{\beta}_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\left(\frac{1}{n} + \frac{(\bar{x} - x_0)^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}\right) \hat{\sigma}^2}}$$

v.a. fulcral se:

- i) $IC_{(1-\alpha)}[E(Y|x_0)]$
- ii) Teste de hipótese para $E(Y|x_0)$

↳ Coeficiente de Determinação

Para além de averiguarmos se o valor esperado da variável resposta (Y) depende linearmente da variável explicativa (x), é crucial verificar se a reta de regressão se ajusta bem ao conjunto de dados. Para o efeito, cingimo-nos a calcular o coeficiente de determinação.

O coeficiente de determinação é definido por

$$\begin{aligned} r^2 &= \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ &= 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ &= \frac{(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y})^2}{(\sum_{i=1}^n x_i^2 - n\bar{x}^2) \times (\sum_{i=1}^n y_i^2 - n\bar{y}^2)} \end{aligned}$$

e dá uma ideia aproximada do ajustamento da reta de regressão aos dados.

- $r^2 \in [0, 1]$
- $r^2 = 0 \rightarrow$ Não há relação linear entre x e y
- $r^2 = 1 \rightarrow$ Todos os pontos estão sobre a reta de regressão ($\hat{y}_i = y_i, i = 1, \dots, n$), pelo que o modelo de RLS explica toda a variação observada na variável resposta Y , logo o modelo de RLS ajustado é PERFEITO.⁷

Em particular $r^2 \times 100\%$ corresponde à percentagem da variação total da variável resposta Y ($\sum_{i=1}^n (y_i - \bar{y})^2$) explicada pela variável regressora x através do modelo de RLS estimado ($\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$).

⁷No caso em que $n = 2$ e $x_1 \neq x_2$, $r^2 = 1$, no entanto, o modelo de RLS pode não fazer qualquer sentido para explicar a relação entre as variáveis resposta e explicativa.

Referências

- [1] Manuel Cabral Morais. *Probabilidades e Estatística: Teoria, Exemplos e Exercícios*. IST Press, 2020. ISBN 978-989-8481-78-8.
- [2] Sheldon M. Ross. *Introduction to Probability and Statistics for Engineers and Scientists*. Academic Press, 2021. ISBN 978-0-12-824346-6.
- [3] F. M. Dekking, C. Kraaikamp, H. P. Lopuhaä, and L. E. Meester. *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer London, 2005. ISBN 978-1-85233-896-1.
- [4] Norman Matloff. *Probability and Statistics for Data Science: Math + R + Data*. CRC Press, 2019. ISBN 978-1138393295.