Universidade Nova de Lisboa
Faculdade de Ciências e Tecnologia

Probabilidades e Estatística

Probabilidades

Ano Lectivo 2011/2012

Nota introdutória

Este documento de apoio à disciplina de Probabilidades e Estatística tem por base apontamentos elaborados pela Professora Doutora Fátima Miguens. Estes foram alvo de uma revisão, com consequente correcção de pequenas gralhas e algumas alterações ao nível da estrutura do texto, durante os anos lectivos 2008/09 e 2009/10, sob a Regência do Professor Doutor Rui Cardoso. A ambos o nosso muito obrigado pelo documento que disponibilizam para o presente ano lectivo.

A leitura destes apontamentos não dispensa a leitura atenta das obras indicadas como referências bibliográficas.

Conteúdo

1 TEORIA DAS PROBABILIDADES **5**
- 1.1 Introdução 5
- 1.2 Espaço de resultados e acontecimento 5
- 1.3 Álgebra de acontecimentos 6
- 1.4 Probabilidade e axiomática das probabilidades 7
- 1.5 Probabilidade condicional 9
  - 1.5.1 Teorema da probabilidade total; Teorema de Bayes 10
- 1.6 Independência de acontecimentos 12

2 VARIÁVEIS ALEATÓRIAS **13**
- 2.1 Definição 13
- 2.2 Função de distribuição 15
- 2.3 Variável aleatória discreta 16
- 2.4 Variável aleatória absolutamente contínua 18
- 2.5 Vectores aleatórios 21
  - 2.5.1 Par aleatório discreto 22
  - 2.5.2 Par aleatório contínuo 24

3 MOMENTOS E PARÂMETROS **26**
- 3.1 Valor médio ou esperança matemática 26
  - 3.1.1 Propriedades do valor médio 28
- 3.2 Variância e desvio padrão 29
  - 3.2.1 Propriedades da variância 31
- 3.3 Covariância e coeficiente de correlação 31
  - 3.3.1 Propriedades da covariância 34
  - 3.3.2 Propriedades do coeficiente de correlação 34
- 3.4 Outros valores esperados sobre um par aleatório 35
- 3.5 Outros parâmetros de localização de uma variável aleatória 35
  - 3.5.1 A mediana 35
  - 3.5.2 O quantil 36
  - 3.5.3 A moda 36
- 3.6 Outros parâmetros de dispersão de uma variável aleatória 36
  - 3.6.1 O desvio médio 36

4 DISTRIBUIÇÕES IMPORTANTES **37**
- 4.1 Distribuições discretas 37
  - 4.1.1 Distribuição Hipergeométrica 37
  - 4.1.2 Distribuição Binomial 39

4.1.3 Distribuição de Poisson 44
4.1.4 Processo de Poisson 47
4.1.5 Distribuição Binomial Negativa e Distribuição Geométrica 47
4.2 Distribuições contínuas 50
4.2.1 Distribuição Uniforme 50
4.2.2 Distribuição Exponencial 50
4.2.3 Distribuição Normal 53
4.2.4 Distribuição Qui-quadrado 57
4.2.5 Distribuição t-student 58
4.2.6 Distribuição F de Fisher 59

5 TEOREMA LIMITE CENTRAL 61
5.1 Aplicações particulares do T.L.C. 63
5.1.1 Distribuição binomial 63
5.1.2 Distribuição de Poisson 65

Lista de Figuras

4.1 Função densidade da distribuição Uniforme 50
4.2 Exemplos de funções densidade da distribuição exponencial 51
4.3 Exemplos de funções densidade da distribuição Normal 54
4.4 Função distribuição de Z 54
4.5 Consequência da simetria da densidade de Z para a sua função de distribuição 55
4.6 Função densidade de $Y \sim \chi^2_n$. 58
4.7 Função densidade de $T \sim t_n$. 58
4.8 Função densidade de $X \sim F_{m,n}$. 60

5.1 Exemplificação do Teorema Limite Central 62
5.2 Ilustração da aproximação da dist. Binomial pela dist. Normal para $n = 2$ e $n = 10$. 64
5.3 Ilustração da aproximação da dist. Binomial pela dist. Normal para $n = 30$ 65
5.4 Ilustração da aproximação da dist. Binomial pela dist. Normal para $n = 50$ 65
5.5 Ilustração da aproximação da dist. Poisson pela dist. Normal para $\lambda = 1, 5, 10$ e 20 . . 66

Capítulo 1

TEORIA DAS PROBABILIDADES

1.1 Introdução

A **teoria das probabilidades** tem como objectivo a formulação de modelos de fenómenos em que intervém o acaso.

Fenómenos aleatórios são os fenómenos sujeitos à influência do acaso e, como tal, não controláveis pelo homem.

Definição 1.1 *Dá-se o nome de* experiência aleatória *a todo o procedimento cujo resultado é imprevisível (fenómeno aleatório), mas que*

- *podemos repetir um grande número de vezes nas mesmas condições (ou em condições muito semelhantes);*
- *com a longa repetição da experiência os resultados patenteiam uma regularidade de observação;*
- *conhecemos o conjunto de todos os resultados possíveis de ocorrerem.*

Exemplo 1.1 *E1 Lançamento de um dado e observação do número de pintas da face que fica virada para cima;*

E2 Número de peças defeituosas fabricadas por uma máquina durante 24 horas;

E3 Tempo de vida de uma pessoa, em anos;

E4 Número de peças fabricadas por uma máquina até se observarem 10 defeituosas;

E5 Tipo de aproveitamento (qualitativo) de um aluno da FCT/UNL.

1.2 Espaço de resultados e acontecimento

Definição 1.2 *Dá-se o nome de* espaço de resultados *ou* universo *de uma experiência aleatória (e representa-se por Ω), ao conjunto de todos os possíveis resultados dessa experiência.*

Exemplo 1.2

E1 $\Omega = \{1, 2, 3, 4, 5, 6\}$*;*

E2 $\Omega = \{0, 1, 2, \ldots, N\}$*, sendo* N *o total de peças produzidas em 24 horas;*

E3 $\Omega = \{1, 2, \ldots\} = \mathbb{N}$*;*

E4 $\Omega = \{10, 11, 12, \ldots\}$*;*

E5 $\Omega = \{Faltou, Desistiu, Aprovado, Reprovado\}$*.*

Definição 1.3 *Dá-se o nome de* acontecimento aleatório *a qualquer subconjunto do espaço de resultados,* Ω*.*

Exemplo 1.3

E1 A=Saída de face com um número par de pintas *corresponde a* $A = \{2, 4, 6\}$*;*

E3 B=Uma pessoa viver até aos 60 anos *corresponde a* $B = \{1, 2, \ldots, 60\}$*;*

E5 C=Alunos sem nota *corresponde a* $C = \{Faltou, Desistiu\}$*.*

Diz-se que um *acontecimento se realiza* sempre que o resultado da experiência aleatória é um elemento que pertence ao acontecimento.

1.3 Álgebra de acontecimentos

1. A é *sub-acontecimento* de B e escreve-se $A \subset B$, se e só se a realização de A implica a realização de B;

2. Dado o acontecimento A, chama-se *acontecimento complementar de* A e escreve-se $\overline{A}$, ao acontecimento constituído pelos elementos de Ω que não estão em A;

3. Dados os acontecimentos A e B, dá-se o nome de *união de* A *com* B ao acontecimento que consiste na realização de pelo menos um deles e representa-se por $A \cup B$;

4. *Intersecção de* A *com* B é o acontecimento que se realiza se, e só se, se realizam em simultâneo os acontecimentos A e B. Representa-se por $A \cap B$;

5. Chama-se *diferença dos acontecimentos* A *e* B ao acontecimento $A - B = A \cap \overline{B}$, ou seja ao acontecimento que se realiza se, e só se, A se realiza mas não se realiza B.

- A Ω chamamos *acontecimento certo*;

- A $\emptyset$ dá-se o nome de *acontecimento impossível*;

- Ao acontecimento que é formado por um único elemento chamamos *acontecimento elementar*;

- Dois acontecimentos A e B dizem-se *mutuamente exclusivos* ou *disjuntos* se não têm elementos em comum, ou seja, se $A \cap B = \emptyset$.

Exemplo 1.4 *Considere-se a experiência aleatória que consiste no lançamento de dois dados e registo da soma das pintas das duas faces viradas para cima.*

Espaço de resultados: $\Omega = \{2, 3, 4, \ldots, 10, 11, 12\}$

Definam-se os acontecimentos:

A - A soma das pintas é um número par*;*

B - A soma das pintas é um número menor que 5*.*

$A = \{2, 4, 6, 8, 10, 12\}$; $B = \{2, 3, 4\}$;

Ocorrer pelo menos um dos acontecimentos $A \cup B = \{2, 3, 4, 6, 8, 10, 12\}$;

Ocorrer A e B $A \cap B = \{2, 4\}$;

Ocorrer A mas não B $A - B = \{6, 8, 10, 12\}$;

Não ocorrer A $\overline{A} = \{3, 5, 7, 9, 11\}$.

1.4 Probabilidade e axiomática das probabilidades

Definição 1.4 *A* probabilidade *mede a possibilidade de ocorrência de um determinado acontecimento.*

Conceito Clássico de Probabilidade

Definição 1.5 (Lei de Laplace) *Se uma experiência aleatória tem N resultados mutuamente exclusivos e* igualmente prováveis (ou possíveis)*e, se desses resultados, N_A têm um certo atributo A, então a* probabilidade de A *é dada por $\frac{N_A}{N}$. Usualmente, enuncia-se esta lei por:*

$$P(A) = \frac{N_A}{N} = \frac{n^o \text{ de resultados favoráveis}}{n^o \text{ de casos possíveis}}$$

Exemplo 1.5

$$P(A) = \frac{6}{11} \quad P(A \cap B) = \frac{2}{11} \quad P\left(\overline{A}\right) = \frac{5}{11}$$

Conceito Frequencista de Probabilidade

E se os resultados da experiência não são igualmente prováveis?...

Definição 1.6 *A probabilidade do acontecimento A é avaliada através de informação existente sobre A, sendo igual à razão entre o número de vezes em que se observou a sua realização (n_A) e o número de vezes (n) em que se efectuou a experiência aleatória, isto é,*

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

Conceito Axiomático de Probabilidade

Segundo este conceito, a probabilidade é uma função de acontecimentos e, quando aplicada a um acontecimento A expressa a “credibilidade”que se atribui à realização deste acontecimento.

Evidentemente que a função *probabilidade* deve satisfazer determinados requisitos que permitem a consistência entre este conceito de probabilidade e outros conceitos, como por exemplo os dois conceitos que apresentámos anteriormente.

Consideremos o espaço de resultados Ω finito e $\mathcal{P}(\Omega)$ o conjunto das partes de Ω.

Definição 1.7 *Chama-se* probabilidade P*, a uma aplicação do conjunto das partes de* Ω *no intervalo* $[0,1]$*, i.e.,*

$$P: \quad \mathcal{P}(\Omega) \rightarrow [0,1]$$

satisfazendo os seguintes axiomas:

A1 $P(A) \geq 0 \quad \forall A \in \mathcal{P}(\Omega)$*;*

A2 $P(\Omega) = 1$*;*

A3 $P(A \cup B) = P(A) + P(B)$*, se* $A \cap B = \emptyset$*, com* $A, B \in \mathcal{P}(\Omega)$*.*

Quando Ω *é infinito, é necessário considerar a generalização do axioma A3 ao caso de infinitos acontecimentos. Tem-se então o axioma A3 substituído pelo axioma*

*A3** $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$*, se* $A_i \cap A_j = \emptyset, \quad i \neq j$*, com* $A_1, A_2, \ldots \in \mathcal{P}(\Omega)$*.*

Consequências da axiomática das probabilidades

1. $P\left(\overline{A}\right) = 1 - P(A)$;
2. $P(\emptyset) = 0$;
3. $A \subset B \Rightarrow P(A) \leq P(B)$;
4. $P(A) \leq 1$;
5. $P(A - B) = P(A) - P(A \cap B)$;
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;
7. $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i \neq j} P(A_i \cap A_j) + \sum_{i \neq j \neq k} P(A_i \cap A_j \cap A_k) + \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$

Definição 1.8 *Dois acontecimentos A e B dizem-se* incompatíveis *se* $P(A \cap B) = 0$*.*

Exemplo 1.6

$$P(A \cup B) = \frac{6}{11} + \frac{3}{11} - \frac{2}{11} = \frac{7}{11} \qquad P(A - B) = \frac{6}{11} - \frac{2}{11} = \frac{4}{11} \qquad P\left(\overline{A}\right) = 1 - \frac{6}{11} = \frac{5}{11}$$

Outros Exemplos

1. Suponha que inadvertidamente se misturaram duas lâmpadas defeituosas com 4 lâmpadas boas.
   (a) Se retirarmos ao acaso duas lâmpadas, qual a probabilidade de ambas serem boas?
   (b) Se retirarmos ao acaso três lâmpadas, qual a probabilidade de todas serem boas?
   (c) Retirámos uma lâmpada e verificámos que é boa. Qual a probabilidade de serem boas as duas lâmpadas que a seguir viermos a extrair?

2. Na próxima jornada de futebol as equipas dos Craques e dos Invencíveis vão jogar jogos separados. Os Craques têm uma probabilidade de 0.4 de vencer o seu adversário enquanto que os Invencíveis poderão vencer o seu oponente com 0.3 de probabilidade. Um jogador do totobola sabe que ambas as equipas vencerão com probabilidade 0.1.

   Qual a probabilidade de, no fim da jornada,

   (a) Alguma destas equipas ter ganho?

   (b) Nenhuma delas ter ganho?

   (c) Somente os Invencíveis ganharem o respectivo desafio?

3. Alguns alunos de uma determinada escola praticam uma ou mais de 3 modalidades desportivas, nomeadamente, futebol, basquetebol e andebol. São conhecidas as seguintes proporções:

   - 30% praticam futebol;
   - 20% praticam basquetebol;
   - 20% praticam andebol;
   - 5% praticam futebol e basquetebol;
   - 10% praticam futebol e andebol;
   - 5% praticam basquetebol e andebol;
   - 2% praticam todas estas modalidades.

   (a) se escolhermos um aluno ao acaso, qual a probabilidade de ser:

      i. Um jogador de futebol ou de andebol?

      ii. Apenas jogador de futebol?

      iii. Um atleta?

   (b) Se escolhermos um aluno atleta, qual a probabilidade de ser:

      i. Apenas jogador de futebol?

      ii. Um jogador de futebol ou de andebol?

1.5 Probabilidade condicional

Num supermercado são vendidas embalagens de café de três marcas A, B e C. Algumas embalagens estão fora de prazo de validade (F). No quadro que se segue, apresentam-se as proporções de embalagens de café, segundo a marca e o estado de validade.

	A	B	C	Total
F	0.02	0.10	0.00	0.12
$\overline{F}$	0.38	0.40	0.10	0.88
Total	0.40	0.50	0.10	1.00

Qual a probabilidade de uma embalagem de café, escolhida ao acaso, estar fora do prazo de validade?

$P\left(F\right) = 0.12$

De entre as embalagens de café da marca B, qual a probabilidade de escolher uma fora do prazo de validade?

Há que tomar em atenção que queremos a proporção de embalagens fora do prazo de validade, considerando apenas o conjunto (ou sub-população) das embalagens da marca B, logo a resposta será

$$\frac{0.10}{0.50} = 0.2$$

Pretendeu-se conhecer a probabilidade do acontecimento "Embalagem de café fora do prazo de validade", sabendo que o acontecimento "Embalagem de café da marca B" é um acontecimento certo. De outro modo, o espaço de resultados deixa de ser o conjunto de todas as embalagens e passa a ser só o conjunto das embalagens da marca B.

Este novo acontecimento deve representar-se por

$$F\,|B$$

em que a notação $\cdot\,|B$ serve para informar que o espaço de resultados foi alterado, isto é, deixa de ser $\Omega = \{\text{Todas as embalagens de café}\}$ e passa agora a ser $B = \{\text{Embalagens da marca B}\}$.

Repare também que o quociente apresentado no cálculo da probabilidade, corresponde a

$$\frac{P\,(F \cap B)}{P\,(B)}$$

Definição 1.9 *Sejam A e B acontecimentos em que $P\,(B) > 0$. A probabilidade de se realizar A sabendo que (dado que, se) B se realizou, designa-se por* probabilidade condicional *de A dado B, e define-se por*

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

Exercício 1.1 *Provar que se B é um acontecimento tal que $P\,(B) > 0$, então $P\,(\cdot\,|B\,)$ é uma probabilidade sobre Ω.*

Teorema 1.1 (Teorema da probabilidade composta) *Se $P\,(A) > 0$ e $P\,(B) > 0$, tem-se*

$$P\,(A \cap B) = P\,(A\,|B\,)\,P\,(B) = P\,(B\,|A\,)\,P\,(A)$$

Exercício 1.2 *Mostrar que, se $P\,(A) > 0$, $P\,(B) > 0$ e $P\,(C) > 0$, então*
$P\,(A \cap B \cap C) = P\,(C\,|A \cap B\,)\,P\,(B\,|A\,)\,P\,(A)$

1.5.1 Teorema da probabilidade total; Teorema de Bayes

Definição 1.10 *Os acontecimentos $\{B_1, B_2, \ldots, B_n\}$ constituem uma partição do espaço de resultados Ω se:*

- *Forem todos disjuntos, isto é, $B_i \cap B_j = \emptyset \quad i \neq j,\ i, j = 1, \ldots, n$;*
- *A união de todos é igual a Ω, ou seja, $\displaystyle\bigcup_{i=1}^{n} B_i = \Omega$.*

Teorema 1.2 (Teorema da probabilidade total) *Se $\{B_1, B_2, \ldots, B_n\}$ é uma partição do espaço de resultados Ω, com $P(B_i) > 0$, para $i = 1, 2, \ldots, n$, dado um qualquer acontecimento A, tem-se*

$$P(A) = P(A|B_1)\,P(B_1) + P(A|B_2)\,P(B_2) + \ldots + P(A|B_n)\,P(B_n)$$

Demonstração: Se $\{B_1, B_2, \ldots, B_n\}$ é uma partição de Ω, então $B_1 \cup B_2 \cup \ldots \cup B_n = \Omega$ e $B_i \cap B_j = \emptyset$, $\forall i \neq j$. Podemos então desmembrar o acontecimento A em

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n)$$

Como os acontecimentos $(A \cap B_1), (A \cap B_2), \ldots (A \cap B_n)$ são todos disjuntos, pelo axioma A3,

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n)$$

Mas

$$P(A \cap B_i) = P(A|B_i)\,P(B_i), \quad \forall i = 1, 2, \ldots, n$$

logo

$$P(A) = P(A|B_1)\,P(B_1) + P(A|B_2)\,P(B_2) + \ldots + P(A|B_n)\,P(B_n)$$

Teorema 1.3 (Teorema de Bayes) *Se $\{B_1, B_2, \ldots, B_n\}$ é uma partição do espaço de resultados Ω, com $P(B_i) > 0$, para $i = 1, 2, \ldots, n$, dado um qualquer acontecimento A, com $P(A) > 0$, tem-se*

$$P(B_i|A) = \frac{P(A|B_i)\,P(B_i)}{\sum_{i=1}^{n} P(A|B_i)\,P(B_i)}$$

Demonstração: Pelo teorema da probabilidade total e pela definição de probabilidade condicional,

$$P(B_i|A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(A|B_i)\,P(B_i)}{\sum_{i=1}^{n} P(A|B_i)\,P(B_i)}$$

Exemplo 1.7 *Certa empresa obtém os seus fornecimentos de três origens B_1, B_2 e B_3, cujas proporções de fornecimento e percentagens de fornecimento de lotes defeituosos são referidos no seguinte quadro:*

	% Fornecimento	*% Defeituosos*
B_1	*0.45*	*5%*
B_2	*0.25*	*7%*
B_3	*0.30*	*10%*

$\{B_1, B_2, B_3\}$ constitui uma partição de Ω. Seja D o acontecimento "lote ser defeituoso".
Probabilidades conhecidas:
$P(B_1) = 0.45,\ P(B_2) = 0.25,\ P(B_3) = 0.30$
$P(D|B_1) = 0.05,\ P(D|B_2) = 0.07,\ P(D|B_3) = 0.10.$

- *Qual a probabilidade de se encontrar um lote defeituoso, de entre o fornecimento total?*
  $P(D) = P(D|B_1)\,P(B_1) + P(D|B_2)\,P(B_2) + P(D|B_3)\,P(B_3) = 0.07$

- *Um lote é defeituoso e não se sabe a sua proveniência. Qual a origem a que se deve apresentar a reclamação?*
  $P(B_1|D) = \dfrac{P(D|B_1)P(B_1)}{P(D)} = 0.32$
  $P(B_2|D) = \dfrac{P(D|B_2)P(B_2)}{P(D)} = 0.25$
  $P(B_3|D) = 1 - P(B_1|D) - P(B_2|D) = 0.43$
  Devemos reclamar a origem B_3.

1.6 Independência de acontecimentos

Definição 1.11 *Dados dois acontecimentos A e B, com $P(B) > 0$, se o conhecimento de que B ocorreu não influencia a probabilidade de que A vir a acontecer, então*

$$P(A|B) = P(A)$$

e dizemos que os acontecimentos são ***independentes****.*

Mas,

$$\begin{aligned} & P(A|B) = P(A) \Leftrightarrow \\ \Leftrightarrow\quad & \frac{P(A\cap B)}{P(B)} = P(A) \Leftrightarrow \\ \Leftrightarrow\quad & P(A\cap B) = P(A)\,P(B) \end{aligned}$$

Definição 1.12 *Dois acontecimentos A e B dizem-se* ***independentes*** *se, e só se*

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

Teorema 1.4 *Se A e B são acontecimentos independentes, também o são A e $\overline{B}$, $\overline{A}$ e B e ainda $\overline{A}$ e $\overline{B}$.*

Exemplo 1.8 *A probabilidade de um atirador acertar no alvo, em cada tiro, é de 0.6, independentemente do tiro. Qual a probabilidade de:*

a) *Serem necessários exactamente 10 tiros para acertar uma vez?*

b) *Em três tiros acertar uma vez?*

c) *Acertar pela terceira vez ao quinto tiro?*

d) *Necessitar de, pelo menos 4 tiros, para acertar duas vezes?*

Capítulo 2

VARIÁVEIS ALEATÓRIAS

2.1 Definição

Numa experiência aleatória os elementos do espaço de resultados Ω podem ser números reais. Por exemplo, o registo do nível de poluição a uma dada hora numa zona urbana, o registo da temperatura máxima diária, o total de pontos obtidos com o lançamento de dois dados, etc. Já o mesmo não se passa se quisermos registar em cada dia, se o céu está muito, pouco ou não nublado, ou se uma equipa de futebol ganha, perde ou empata um jogo, etc.

Quando Ω não é constituído por elementos de carácter quantitativo, a necessidade de aplicação de procedimentos estatísticos obriga a uma atribuição de valores numéricos a cada elemento ω de Ω. Essa atribuição pode ser arbitrária ou não.

Exemplo 2.1 *Se considerarmos a experiência aleatória de lançamento ao ar de uma moeda e registo da face que fica virada para cima, o espaço de resultados é $\Omega = \{Cara, Coroa\}$ a que podemos atribuir*

ω	$X(\omega)$
Cara	*1*
Coroa	*0*

É evidente que podemos definir diversas correspondências sobre o mesmo universo Ω.

Exemplo 2.2 *Consideremos uma população de empresas, das quais se escolhe uma ao acaso. Se existirem m empresas, então o universo é $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$ onde ω_i representa a empresa i.*

Sobre este universo de empresas podemos definir várias correspondências (ou variáveis aleatórias):
$\omega \to X(\omega)$, *sendo $X(\omega)$ o número de empregados da empresa ω;*
$\omega \to Y(\omega)$, *sendo $Y(\omega)$ o valor anual dos impostos cobrados à empresa ω;*
$\omega \to Z(\omega)$, *sendo $Z(\omega)$ o volume de vendas anual da empresa ω;*
e muitas mais.

Pensando só numa correspondência sobre Ω, se associarmos a cada elemento $\omega \in \Omega$ um número real $X(\omega)$, estamos a definir uma função $X : \Omega \to \mathbb{R}$.

Sendo A um acontecimento, chamamos *imagem de A por X*, ao conjunto de valores que X assume para os elementos ω de A, isto é,

$$X(A) = \{X(\omega) : \omega \in A\}$$

Por outro lado, a cada subconjunto $E \subset \mathbb{R}$, pode-se fazer corresponder o subconjunto $X^{-1}(E)$ formado por todos os elementos $\omega \in \Omega$ tais que $X(\omega) \in E$,

$$X^{-1}(E) = \{\omega : \ X(\omega) \in E\}$$

$X^{-1}(E)$ denomina-se *imagem inversa de E por X.*

Exemplo 2.3 *No lançamento de dois dados, interessa-nos saber o valor da soma das pintas das faces viradas para cima. O espaço de resultados será*

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

e define-se a aplicação

$$X(i, j) = i + j$$

Se $A = \{(1,1), (1,2), (2,1)\}$, a imagem de A por X é $X(A) = \{2, 3\}$.

Para o subconjunto real $E_1 = \{3, 11\}$, a imagem inversa de E_1 por X é o acontecimento $X^{-1}(E_1) = \{(1,2), (2,1), (5,6), (6,5)\}$.

Para o subconjunto real $E_2 = [3, 4]$, a imagem inversa de E_2 por X é o acontecimento $X^{-1}(E_2) = \{(1,2), (2,1), (1,3), (3,1), (2,2)\}$.

Para o subconjunto real $E_3 = [2, +\infty[$, a imagem inversa de E_3 por X é o acontecimento $X^{-1}(E_3) = \Omega$.

Para o subconjunto real $E_4 =]-\infty, 0.7]$, a imagem inversa de E_4 por X é o acontecimento $X^{-1}(E_4) = \emptyset$.

Definição 2.1 *Seja Ω um espaço de resultados associado a uma experiência aleatória. Chama-se* ***variável aleatória*** *(abreviadamente,* ***v.a.****) a uma função $X : \Omega \longrightarrow \mathbb{R}$ tal que $A_x = \{\omega \in \Omega : \ X(\omega) \leq x\}$, com $x \in \mathbb{R}$, é um acontecimento.*

Exemplo 2.4 *O espaço de resultados associado ao lançamento de uma moeda por três vezes pode ser definido por:*

$$\Omega = \{(FFF), (FFC), (FCF), (CFF), (FCC), (CFC), (CCF), (CCC)\}$$

Considere-se a seguinte variável aleatória: X = nº de coroas obtidas nos três lançamentos.

Esta v.a. tem como contradomínio o subconjunto $\{0, 1, 2, 3\}$ de $\mathbb{R}$ e, admitindo que a probabilidade de sair coroa em cada lançamento é de $1/3$ e que estes se processam independentemente uns dos outros,

$$\begin{aligned}
P(X=0) &= P(\{(FFF)\}) = \frac{8}{27} \\
P(X=1) &= P(\{(CFF), (FCF), (FFC)\}) = \frac{12}{27} \\
P(X=2) &= P(\{(CCF), (CFC), (FCC)\}) = \frac{6}{27} \\
P(X=3) &= P(\{(CCC)\}) = \frac{1}{27}
\end{aligned}$$

A partir daqui, podemos determinar a probabilidade de muitos outros acontecimentos. Por exemplo, a probabilidade de se observar pelo menos 2 coroas:

$$P(X \geq 2) = P(X=2) + P(X=3) = \frac{7}{27}$$

E a probabilidade de se observar menos de 3 coroas:

$$P(X < 3) = 1 - P(X=3) = \frac{26}{27}$$

Proposição 2.1 *Se $X_1, X_2, \ldots, X_m$ são m v.a.'s e h é uma função contínua de $\mathbb{R}^m$ em $\mathbb{R}$, então $Y = h(X_1, X_2, \ldots, X_m)$ é uma v.a..*

2.2 Função de distribuição

A função distribuição é uma ferramenta de cálculo de probabilidades muito útil e importante.

Definição 2.2 *Designa-se por função distribuição da variável aleatória X, a função F que, para um dado valor x do argumento, associa a probabilidade de X assumir qualquer valor real inferior ou igual a x, ou seja,*

$$F(x) = P(\{\omega : X(\omega) \leq x\}) = P(X^{-1}(]-\infty, x])) = P(X \in]-\infty, x]) = P(X \leq x), \ \forall x \in \mathbb{R}$$

Esta função acumula probabilidades pois, dados dois valores x e y, com $x < y$, dado que $]-\infty, y] =]-\infty, x] \cup]x, y]$, então

$$F(y) = P(X \leq y) = P(X \leq x) + P(x < X \leq y) = F(x) + P(x < X \leq y) \tag{2.2.1}$$

Enunciemos algumas das suas propriedades:

Proposição 2.2 *Qualquer variável aleatória tem uma, e uma só, função distribuição.*

Proposição 2.3 *Uma função distribuição F é uma função real de variável real, não decrescente, contínua à direita e satisfaz $F(-\infty) = 0$ e $F(+\infty) = 1$*

O conhecimento da função de distribuição permite-nos determinar outras probabilidades.

Exemplo 2.5 *Retomemos o exemplo anterior 2.4. Neste caso tem-se*

$$X^{-1}(]-\infty, x]) \begin{cases} \emptyset & \text{se} \quad x < 0 \\ \{FFF\} & \text{se} \quad 0 \leq x < 1 \\ \{FFF, CFF, FCF, FFC\} & \text{se} \quad 1 \leq x < 2 \\ \{FFF, CFF, FCF, FFC, CCF, CFC, FCC\} & \text{se} \quad 2 \leq x < 3 \\ \Omega & \text{se} \quad x \geq 3 \end{cases}$$

logo a função de distribuição é dada por

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{8}{27} & 0 \leq x < 1 \\ \frac{20}{27} & 1 \leq x < 2 \\ \frac{26}{27} & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

2.3 Variável aleatória discreta

Definição 2.3 *Seja X uma variável aleatória e D o conjunto*

$$D = \{a : \ P(X = a) > 0\}$$

(quando muito numerável). A variável aleatória diz-se do tipo ***discreto*** *quando $P(X \in D) = 1$. O conjunto D é designado por* ***suporte*** *da v.a. X.*

Definição 2.4 *Seja X uma v.a. discreta. Chama-se* ***função de probabilidade****(f.p.) de X à função definida no suporte $D = \{a_1, a_2, \ldots\}$ dos valores de X que são observados com probabilidade não nula, isto é, $P(X = a_i) > 0$ e pelas respectivas probabilidades. Uma representação usual para a função de probabilidade da v.a. X, será:*

$$X \left\{ \begin{array}{cccccc} a_1 & a_2 & \ldots & a_i & \ldots \\ P(X = a_1) & P(X = a_2) & \ldots & P(X = a_i) & \ldots \end{array} \right.$$

Propriedades da função de probabilidade

1. $P(X = a_i) > 0, \quad \forall a_i;$
2. $\displaystyle\sum_{i=1}^{\infty} P(X = a_i) = 1$

Cálculo de probabilidades

Dado um qualquer intervalo real, I,

$$P(X \in I) = \sum_{a_i \in D \cap I} P(X = a_i)$$

em particular a sua função de distribuição é determinada por

$$F(x) = P(X \leq x) = \sum_{a_i \leq x} P(X = a_i), \quad \forall\, x \in \mathbb{R} \tag{2.3.2}$$

Exemplo 2.6 *Retomemos o exemplo 2.4. A v.a. X-nº de coroas observadas nos três lançamentos, é uma v.a. discreta e a sua função de probabilidade será:*

$$X \left\{ \begin{array}{cccc} 0 & 1 & 2 & 3 \\ 8/27 & 12/27 & 6/27 & 1/27 \end{array} \right.$$

e

$$\begin{aligned} P(X < 2) &= P(X = 0) + P(X = 1) = 20/27 \\ P(0.7 \leq X < 2.6) &= P(X = 1) + P(X = 2) = 18/27 \end{aligned}$$

Quanto à função de distribuição, apesar de já ter sido deduzida no exemplo 2.5, ilustramos a sua determinação de acordo com a equação 2.3.2

$$\begin{array}{rl} \textit{Se } x < 0 & , \quad F(x) = P(X \leq x) = P(\emptyset) = 0 \\ \textit{Se } 0 \leq x < 1 & , \quad F(x) = P(X \leq x) = P(X = 0) = 8/27 \\ \textit{Se } 1 \leq x < 2 & , \quad F(x) = P(X \leq x) = P(X = 0) + P(X = 1) = 20/27 \end{array}$$

Para os dois casos que restam, usamos para efeitos de exemplificação, a propriedade de acumulação de probabilidades da função distribuição (equação 2.2.1).

$$\begin{aligned} \textit{Se } 2 \leq x < 3 \quad &, \quad F(x) = P(X \leq x) = P(X \leq 1) + P(1 < X \leq x) = 20/27 + P(X = 2) = 26/27 \\ \textit{Se } x \geq 3 \quad &, \quad F(x) = P(X \leq x) = P(X \leq 2) + P(2 < X \leq x) = 26/27 + P(X = 3) = 1 \end{aligned}$$

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{8}{27} & 0 \leq x < 1 \\ \frac{20}{27} & 1 \leq x < 2 \\ \frac{26}{27} & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

Como podemos constatar pelo exemplo anterior, a função de distribuição de uma variável discreta, possui, para além das propriedades já enunciadas, outras que importa realçar:

- trata-se de uma função em "escada";
- os valores reais onde ocorrem discontinuidades ("saltos") são os valores observáveis da variável aleatória;
- a amplitude do "salto" num ponto de descontinuidade a corresponde à probabilidade $P(X = a)$ porque
$$P(X = a) = P(X \leq a) - P(X < a) = F(a) - \lim_{x \to a^-} F(x)$$

Estas propriedades permitem-nos determinar a função de probabilidade de uma variável aleatória discreta partindo da correspondente função de distribuição.

Exemplo 2.7 *Seja X uma variável aleatória com função distribuição*

$$F(x) = \begin{cases} 0 & x < -0.6 \\ 0.2 & -0.6 \leq x < 2 \\ 0.8 & 2 \leq x < 5.7 \\ 1 & x \geq 5.7 \end{cases}$$

Reconhecemos que esta função é uma função em escada pelo que, a variável aleatória X é do tipo discreto. Os saltos ocorrem nos valores reais -0.6, 2 e 5.7 sendo estes os valores observáveis de X. A amplitude dos saltos correspondem à probabilidade de observação destes valores, ou seja

$$\begin{aligned} P(X = -0.6) = F(-0.6) - \lim_{x \to -0.6^-} F(x) = 0.2 - 0 = 0.2 \\ P(X = 2) = F(2) - \lim_{x \to 2^-} F(x) = 0.8 - 0.2 = 0.6 \\ P(X = 5.7) = F(5.7) - \lim_{x \to 5.7^-} F(x) = 1 - 0.8 = 0.2 \end{aligned}$$

e portanto, a função de probabilidade da v.a. X é:

$$X \left\{ \begin{array}{ccc} -0.6 & 2 & 5.7 \\ 0.2 & 0.6 & 0.2 \end{array} \right.$$

Exemplo 2.8 *Voltando ao exemplo 2.6, calculemos algumas probabilidades usando a função de distribuição:*

$$\begin{aligned}
P(X \leq 0.5) &= F(0.5) = 8/27 \\
P(X \geq 1) &= 1 - P(X < 1) = 1 - \lim_{x \to 1^-} F(x) = 1 - 8/27 = 19/27 \\
P(1 < X \leq 2.7) &= P(X \leq 2.7) - P(X \leq 1) = F(2.7) - F(1) = 26/27 - 20/27 = 6/27 \\
P(1 < X < 3) &= P(X < 3) - P(X \leq 1) = \lim_{x \to 3^-} F(x) - F(1) = 26/27 - 20/27 = 6/27 \\
P(0 \leq X < 2) &= P(X < 2) - P(X < 0) = \lim_{x \to 2^-} F(x) - \lim_{x \to 0^-} F(x) = 20/27 - 0 = 20/27
\end{aligned}$$

2.4 Variável aleatória absolutamente contínua

Definição 2.5 *Uma v.a. X diz-se* ***absolutamente contínua*** *se, sendo F_X a sua função de distribuição, existe uma função não negativa f_X, tal que, $\forall x \in \mathbb{R}$,*

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

A função f_X é designada por ***função densidade de probabilidade****, (f.d.p.), ou simplesmente* ***função densidade****.*

Equivalentemente também podemos definir uma v.a. absolutamente contínua da seguinte forma:

Definição 2.6 *Uma v.a. X diz-se* ***absolutamente contínua*** *se o conjunto*

$$D = \{a: \ P(X = a) > 0\} = \emptyset \tag{2.4.3}$$

e existe uma função não negativa f_X (função densidade), tal que, para qualquer intervalo I real,

$$P(X \in I) = \int_I f_X(x)\ dx \tag{2.4.4}$$

Ao conjunto de valores reais $S = \{x \in \mathbb{R}: \ f_X(x) > 0\}$ é dado o nome de ***suporte*** *de v.a. X.*

Propriedades da função densidade de probabilidade

1. $f_X(x) \geq 0, \quad \forall x \in \mathbb{R}$;
2. $\displaystyle\int_{-\infty}^{+\infty} f_X(x)\ dx = 1$

Cálculo de probabilidades

Nota 1 Dado um qualquer intervalo real, I,

$$P(X \in I) = \int_I f_X(x)\ dx$$

Como se trata do integral de uma função não negativa e é sempre convergente, então a $P(X \in I)$, corresponde ao valor da área entre o eixo das abcissas e o gráfico da função f_X no intervalo I considerado.

Nota 2 Se X é uma v.a. contínua, a função de distribuição de uma variável contínua, para além de ser contínua à direita, é também contínua à esquerda. Vejamos porquê:

$$P(X \leq a) = F(a)$$

Mas, pela caracterização 2.4.3 de v.a. contínua (i.e. $\mathbb{P}(X = a) = 0,\ \forall a \in \mathbb{R}$)

$$P(X \leq a) = P(X < a) + \underbrace{P(X = a)}_{=0} = P(X < a)$$

Definindo

$$P(X < a) = \lim_{x \to a^-} F(x) = F\left(a^-\right)$$

concluímos que

$$P(X \leq a) = P(X < a) \Leftrightarrow F(a) = F\left(a^-\right), \quad \forall\, a \in \mathbb{R}$$

ou seja, a função distribuição de X é contínua à esquerda em $\mathbb{R}$.

Nota 3 Se o intervalo I for (com $a < b$) $I = [a, b]$ ou $I =]a, b]$ ou $I = [a, b[$ ou ainda $I =]a, b[$, o valor da sua probabilidade é sempre igual, ou seja,

$$P(X \in I) = \int_a^b f_X(x)\ dx$$

Nota 4 Se X é v.a. contínua, pela definição 2.4.4,

$$\begin{aligned}\lim_{h \to 0} \frac{P(a < X \leq a+h)}{h} &= \lim_{h \to 0} \frac{P(X \leq a+h) - P(X \leq a)}{h} = \lim_{h \to 0} \frac{F(a+h) - F(a)}{h} = \\ &= F'(a) = f(a).\end{aligned}$$

Assim,

$$P(a < X \leq a+h) \approx h\, f(a), \quad h \to 0$$

Nota 5 Se X é v.a. contínua com função densidade f e função distribuição F,

$$f(x) = \begin{cases} F'(x) & \text{nos pontos onde existe derivada} \\ 0, & \text{nos outros pontos} \end{cases}$$

Exemplo 2.9 *Suponhamos que o tempo de vida (em horas) de uma determinada marca de pilhas, X, é uma v.a. com função densidade de probabilidade:*

$$f_X(x) = \begin{cases} c/x^2 & x > 100 \\ 0 & \textit{restantes valores de } x \end{cases}$$

a) c pode ter um valor qualquer?

1. *Como* $f_X(x) \geq 0, \quad \forall x \in \mathbb{R} \Rightarrow c \geq 0$
2. *Como* $\displaystyle\int_{-\infty}^{+\infty} f_X(x)\ dx = 1$, *então*

$$\int_{-\infty}^{+\infty} f_X(x)\ dx = c \int_{100}^{+\infty} \frac{1}{x^2}\ dx = c\left[-\frac{1}{x}\right]_{100}^{+\infty} = \frac{c}{100} = 1 \Rightarrow c = 100$$

b) Qual a é a correspondente função de distribuição? Ora, para $x \leq 100$, temos

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt = \int_{-\infty}^{x} 0\, dt = 0$$

Para $x > 100$, vem

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt = \int_{-\infty}^{100} 0\, dt + \int_{100}^{x} \frac{100}{t^2}\, dt = \left[-\frac{100}{t}\right]_{100}^{x} = 1 - \frac{100}{x}$$

Resumindo,

$$F(x) = \begin{cases} 0 & x < 100 \\ 1 - \frac{100}{x} & x \geq 100 \end{cases}$$

c) Qual a probabilidade de uma pilha durar mais de 500 horas?

$$P\,(X > 500) = \int_{500}^{+\infty} \frac{100}{x^2}\, dx = 100 \left[-\frac{1}{x}\right]_{500}^{+\infty} = 0.2$$

Também podemos usar a função de distribuição para calcular esta probabilidade, pois

$$P\,(X > 500) = 1 - P(X \leq 500) = 1 - F(500) = \frac{100}{500} = 0.2$$

d) Usando unicamente a função de distribuição, quais os valores das seguintes probabilidades?

$$\begin{aligned} P(X \leq 110) &= F(110) = 1 - 100/110 = 1/11 \\ P(X \geq 110) &= 1 - F(110) = 10/11 \\ P(90 < X \leq 120) &= P(X \leq 120) - P(X \leq 90) = F(120) - F(90) = 1 - 110/120 - 0 = 1/6 \\ P(110 < X < 120) &= P(X < 120) - P(X \leq 110) = F(120) - F(110) = 5/66 \end{aligned}$$

2.5 Vectores aleatórios

Exemplo 2.10 *Consideremos uma população de empresas, das quais se escolhe uma ao acaso. Se existirem m empresas, então o universo é $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$ onde ω_i representa a empresa i.*

Sobre este universo de empresas podemos definir várias correspondências (ou variáveis aleatórias):
$\omega \to X(\omega)$*, sendo* $X(\omega)$ *o número de empregados da empresa* ω*;*
$\omega \to W(\omega)$*, sendo* $W(\omega)$ *o encargo total anual em salários da empresa* ω*;*
$\omega \to Y(\omega)$*, sendo* $Y(\omega)$ *o valor anual dos impostos cobrados à empresa* ω*;*
$\omega \to Z(\omega)$*, sendo* $Z(\omega)$ *o volume de vendas anual da empresa* ω*;*
e muitas mais.

Quando se pretende estudar diversas características num mesmo elemento ω do espaço de resultados Ω, faz-se corresponder a cada um desses elementos um ponto $(x_1, x_2, \ldots, x_p)$ de $\mathbb{R}^p$, isto é, considera-se a aplicação

$$w \longrightarrow (X_1(\omega), X_2(\omega), \ldots, X_p(\omega))$$

que substitui o espaço de resultados Ω por $\mathbb{R}^p$.

Definição 2.7 *Se para qualquer ponto $(x_1, x_2, \ldots, x_p) \in \mathbb{R}^p$, o conjunto de Ω,*

$$\{w : X_1(w) \leq x_1, X_2(w) \leq x_2, \ldots, X_p(w) \leq x_p\}$$

é um acontecimento, diz-se que

$$X(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_p(\omega))$$

ou numa notação mais abreviada

$$X = (X_1, X_2, \ldots, X_p)$$

é um ***vector aleatório*** *de dimensão p.*

Quando se pretende estudar apenas duas características numéricas sobre um elemento ω de Ω, necessitamos de um vector aleatório de dimensão 2 dito também *par aleatório*.

Representemos por (X, Y) esse par aleatório.

Classificação de um par aleatório

- (X, Y) é um par aleatório discreto se as v.a.'s X e Y são do tipo discreto.
- (X, Y) é um par aleatório contínuo se as v.a.'s X e Y são do tipo contínuo.
- (X, Y) é um par aleatório misto se uma das v.a.'s X e Y é do tipo discreto e a outra do tipo contínuo.

2.5.1 Par aleatório discreto

Definição 2.8 *Seja (X, Y) um par aleatória discreto. Chama-se* ***função de probabilidade conjunta****(f.p.c) de (X, Y) à função definida pelo conjunto $D = \{(x_1, y_1), (x_2, y_2), \ldots\}$ dos valores de (X, Y) que são observados com probabilidade não nula, e pelas respectivas probabilidades,*

$$p_{ij} = P(X = x_i; Y = y_j), \quad i, j = 1, 2, \ldots$$

verificando as seguintes condições:

1. $0 \leq p_{ij} \leq 1, \quad \forall (x_i, y_j) \in D$
2. $\displaystyle\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{ij} = 1$

Exemplo 2.11 *Seja (X, Y) um par aleatório discreto, representando*

- X*-o nº de carros que chegam a um parque de estacionamento, num dado momento*
- Y*-o nº de lugares vagos neste parque, no mesmo momento*

FUNÇÃO DE PROBABILIDADE CONJUNTA DE (X, Y)

$X \setminus Y$	1	2	3		
1	0.05	0.03	0.02	0.1	$P(X = 1)$
2	0.1	0.06	0.04	0.2	$P(X = 2)$
3	0.15	0.09	0.06	0.3	$P(X = 3)$
4	0.2	0.12	0.08	0.4	$P(X = 4)$
	0.5	0.3	0.2	1	
	$P(Y = 1)$	$P(Y = 2)$	$P(Y = 3)$		

Probabilidade de, num dado momento, chegarem ao parque dois carros e só haver lugar para um

$$P(X = 2; Y = 1) = 0.1$$

Probabilidade de, num dado momento, os lugares vagos serem insuficientes para os automóveis que chegam

$$\begin{aligned} P(X > Y) &= P(X = 2; Y = 1) + P(X = 3; Y = 1) + \\ &+ P(X = 3; Y = 2) + P(X = 4; Y = 1) + \\ &+ P(X = 4; Y = 2) + P(X = 4; Y = 3) = \\ &= 0.1 + 0.15 + 0.09 + 0.2 + 0.12 + 0.08 = 0.74 \end{aligned}$$

Probabilidade de, num certo momento, chegarem três automóveis

$$\begin{aligned} P(X = 3) &= P(X = 3; Y = 1) + P(X = 3; Y = 2) + \\ &+ P(X = 3; Y = 3) = 0.15 + 0.09 + 0.06 = 0.3 \end{aligned}$$

Probabilidade de, num certo momento, haver apenas um lugar vago

$$\begin{aligned} P(Y = 1) &= P(X = 1; Y = 1) + P(X = 2; Y = 1) + \\ &+ P(X = 3; Y = 1) + P(X = 4; Y = 1) = \\ &= 0.05 + 0.1 + 0.15 + 0.2 = 0.5 \end{aligned}$$

Definição 2.9 *Dado um par aleatório discreto (X,Y) é possível definir a* ***função de probabilidade marginal de** X e a* ***função de probabilidade marginal de** Y, do seguinte modo:*

$$
\begin{aligned}
p_{i\cdot} &= P(X=x_i) = \sum_{j=1}^{\infty} P(X=x_i;Y=y_j) = \sum_{j=1}^{\infty} p_{ij}, \quad i=1,2,\ldots \\
p_{\cdot j} &= P(Y=y_j) = \sum_{i=1}^{\infty} P(X=x_i;Y=y_j) = \sum_{i=1}^{\infty} p_{ij}, \quad j=1,2,\ldots
\end{aligned}
$$

Estas duas funções são funções de probabilidade de variáveis aleatórias unidimensionais.

Exemplo 2.12
FUNÇÃO DE PROBABILIDADE MARGINAL DE X

$$
X \left\{ \begin{array}{cccc} 1 & 2 & 3 & 4 \\ 0.1 & 0.2 & 0.3 & 0.4 \end{array} \right.
$$

FUNÇÃO DE PROBABILIDADE MARGINAL DE Y

$$
Y \left\{ \begin{array}{ccc} 1 & 2 & 3 \\ 0.5 & 0.3 & 0.2 \end{array} \right.
$$

Independência das variáveis de um par aleatório discreto

Consideremos os acontecimentos $(X=x_i)$ e $(Y=y_j)$. Pelo que sabemos sobre a independência de acontecimentos, poderemos afirmar que aqueles dois acontecimentos são independentes se, e só se,

$$
\begin{aligned}
P((X=x_i)\cap(Y=y_j)) &= P(X=x_i)\,P(Y=y_j) \\
P(X=x_i;Y=y_j) &= P(X=x_i)\,P(Y=y_j) \\
p_{ij} &= p_{i\cdot}p_{\cdot j}
\end{aligned}
$$

Definição 2.10 *Seja (X,Y) um par aleatório discreto. As v.a.'s X e Y dizem-se* ***independentes*** *(ou diz-se que o par aleatório é independente) se, e só se,*

$$
P(X=x_i;Y=y_j) = P(X=x_i)\,P(Y=y_j), \quad \forall (x_i,y_j)\in D
$$

ou numa notação simplificada, se, e só se,

$$
p_{ij} = p_{i\cdot}p_{\cdot j}, \quad i,j=1,2\ldots
$$

Exemplo 2.13

$$
P(X=1;Y=1)=0.05 \quad e \quad P(X=1)\,P(Y=1)=0.1\times 0.5=0.05
$$

logo $P(X=1;Y=1)=P(X=1)\,P(Y=1)$

$$
P(X=1;Y=2)=0.03 \quad e \quad P(X=1)\,P(Y=2)=0.1\times 0.3=0.03
$$

logo $P(X=1;Y=2)=P(X=1)\,P(Y=2)$

O mesmo se passa para todos os outros valores observáveis do par aleatório (X,Y).

Portanto as v.a.'s X e Y são independentes, ou seja o n^o de carros que chegam ao parque de estacionamento, num dado momento, não tem qualquer relação com o n^o de lugares vagos, no mesmo momento.

2.5.2 Par aleatório contínuo

Definição 2.11 *Um par aleatório* (X, Y) *diz-se* ***absolutamente contínuo*** *(ou* ***contínuo****) se o conjunto*

$$D = \{(x, y) : \ P(X = x; Y = y) > 0\} = \emptyset$$

e existe uma função não negativa $f_{(X,Y)}$*, tal que, para qualquer intervalo* I *de* $\mathbb{R}^2$*,*

$$P((X, Y) \in I) = \int \int_I f_{(X,Y)}(x, y) \ dxdy$$

A função $f_{(X,Y)}$ *é designada por* ***função densidade de probabilidade conjunta****, (f.d.p.c), ou simplesmente* ***função densidade conjunta*** *e deve satisfazer as seguintes condições:*

1. $f_{(X,Y)}(x, y) \geq 0, \quad \forall (x, y) \in \mathbb{R}^2$*;*
2. $\displaystyle\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{(X,Y)}(x, y) \ dxdy = 1$

Interpretação da probabilidade
Como, para um qualquer intervalo I de $\mathbb{R}^2$,

$$P((X, Y) \in I) = \int \int_I f_{(X,Y)}(x, y) \ dxdy$$

e $f_{(X,Y)}$ é uma função não negativa, então a $P((X, Y) \in I)$ corresponde ao valor do volume do espaço delimitado pela função densidade conjunta e pelo intervalo I.

Definição 2.12 *Dado um par aleatório contínuo* (X, Y) *é possível definir a* ***função densidade de probabilidade marginal de*** X *e a* ***função densidade de probabilidade marginal de*** Y*, do seguinte modo:*

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

Estas duas funções são funções densidade de probabilidade de variáveis aleatórias unidimensionais.

Independência das variáveis de um par aleatório contínuo

Definição 2.13 *Seja* (X, Y) *um par aleatório contínuo. As v.a.'s* X *e* Y *dizem-se* ***independentes*** *(ou diz-se que o par aleatório é independente) se, e só se,*

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

Exemplo 2.14 *Os tempos de vida, em centenas de horas, das duas componentes principais de um sistema de controlo são v.a.'s* (X, Y) *com função densidade conjunta*

$$f(x, y) = \begin{cases} cx^2 y & 0 < x < 3, 0 < y < 2 \\ 0 & \text{outros valores de} \, (x, y) \in \mathbb{R}^2 \end{cases}$$

a) Qual o valor de c?

$$\begin{aligned} f_{(X,Y)}(x,y) \geq 0, \quad \forall (x,y) \in \mathbb{R}^2 &\Rightarrow \quad c \geq 0 \\ \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{(X,Y)}(x,y)\,dxdy = 1 &\Leftrightarrow \quad \int_0^3\int_0^2 cx^2y\,dxdy = 1 \Leftrightarrow \\ &\Leftrightarrow \quad c = 1/18 \end{aligned}$$

b) Qual a probabilidade de cada uma das componentes durar mais de 100 horas?

$$P(X>1;Y>1) \quad = \quad \int_1^3\int_1^2 \frac{1}{18}x^2y\,dxdy = \frac{13}{18}$$

c) Qual a probabilidade da 1ª componente durar mais de 100 horas?

$$\begin{aligned} P(X>1) &= \int_1^3 f_X dx = ? \\ f_X(x) &= \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y)\,dy = \int_0^2 \frac{1}{18}x^2y\,dy = \frac{x^2}{9},\ 0<x<3 \\ P(X>1) &= \int_1^3 \frac{x^2}{9}dx = \frac{26}{27} \end{aligned}$$

d) Os tempos de vida das componentes são independentes?

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y)\,dx = \int_0^3 \frac{1}{18}x^2y\,dx = \frac{y}{2},\ 0<y<2 \\ f_Y(y) &= \begin{cases} y/2 & 0<y<2 \\ 0 & o.\ v.\ de\ y \end{cases} \qquad f_X(x) = \begin{cases} x^2/9 & 0<x<3 \\ 0 & o.\ v.\ de\ x \end{cases} \\ f(x,y) &= \begin{cases} \frac{1}{18}x^2y & 0<x<3, 0<y<2 \\ 0 & o.\ v.(x,y) \end{cases} = f_X(x)\,f_Y(y) \end{aligned}$$

e) Com a informação da alínea anterior, a alínea b) poder-se-ia determinar de outro modo,

$$\begin{aligned} P(X>1;Y>1) &= P(X>1)\,P(Y>1) = \frac{26}{27} \times \frac{3}{4} = \frac{13}{18} \\ P(Y>1) &= \int_1^2 f_Y(x)\,dy = \int_1^2 \frac{y}{2}dy = \frac{3}{4} \end{aligned}$$

Capítulo 3

MOMENTOS E PARÂMETROS

3.1 Valor médio ou esperança matemática

Exemplo 3.1 *Uma empresa de aluguer de aviões sabe que a procura diária X tem carácter aleatório e estima a sua função de probabilidade por*

$$X \begin{cases} 0 & 1 & 2 & 3 \\ 0.25 & 0.35 & 0.30 & 0.10 \end{cases}$$

Qual o número médio de aviões procurados diariamente?

A resposta será

$$0 \times 0.25 + 1 \times 0.35 + 2 \times 0.30 + 3 \times 0.10 = 1.25 \text{ aviões,}$$

*a que dá o nome de **valor médio** ou **esperança matemática** de X e que se representa por $E(X)$.*

Devido à segunda designação, a quantidade que acabámos de calcular também pode ser enunciada por:

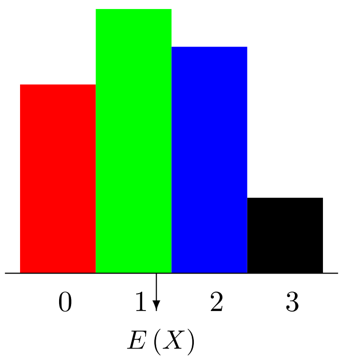
"O número esperado de aviões procurados diariamente é de 1.25.

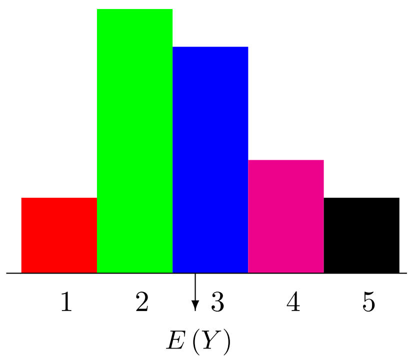
Admitamos que outra empresa de aluguer de aviões apresenta a seguinte função de probabilidade para a procura diária Y,

$$Y \begin{cases} 1 & 2 & 3 & 4 & 5 \\ 0.10 & 0.35 & 0.30 & 0.15 & 0.10 \end{cases}$$

O número médio de aviões procurados por dia é de

$$E(Y) = 1 \times 0.10 + 2 \times 0.35 + 3 \times 0.30 + 4 \times 0.15 + 5 \times 0.10 = 2.8 \text{ aviões.}$$





O valor médio é uma **medida de localização** (ou um **parâmetro de localização**) da v.a. a que se aplica.

Definição 3.1 *Se X é uma v.a., define-se o **valor médio** ou **esperança matemática** de X, por*

a) $E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i)$, *caso X seja uma v.a. discreta com valores no suporte* $D = \{x_1, x_2, \ldots\}$ *e* $\sum_{i=1}^{\infty} |x_i| P(X = x_i) < +\infty$;

b) $E(X) = \int_{-\infty}^{+\infty} x f_X(x)\, dx$, *caso X seja uma v.a. contínua com função densidade de probabilidade* f_X *e* $\int_{-\infty}^{+\infty} |x| f_X(x)\ dx < +\infty$.

Exemplo 3.2 *Suponha que o seu médico lhe aconselha que faça uma dieta para emagrecimento, durante 2 semanas. Considerando a sua estrutura física, pressupõe que o peso (em kg) que vai perder se situa, com igual probabilidade, entre 2 e 4 kg. Quantos quilos espera perder nas duas semanas?*

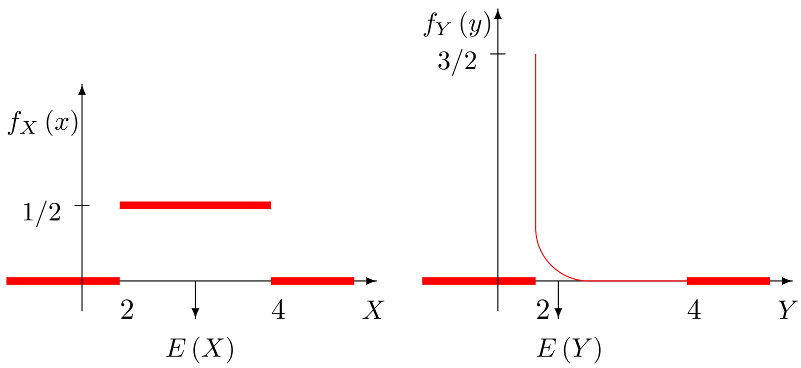
$$f_X(x) = \begin{cases} 1/2 & 2 \leq x \leq 4 \\ 0 & o.v.\ de\ x \end{cases}$$

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)\ dx = \int_2^4 x \frac{1}{2}\ dx = 3\ kg$$

Suponha que o médico lhe propõe outro tipo de dieta, informando-o de que a distribuição do peso é diferente e, é bem descrito pela função densidade de probabilidade

$$f_Y(y) = \begin{cases} \frac{3}{8}(y-4)^2 & 2 \leq y \leq 4 \\ 0 & o.v.\ de\ y \end{cases}$$

$$E(Y) = \int_{-\infty}^{+\infty} y f_Y(y)\ dy = \int_2^4 y \frac{3}{8}(y-4)^2\ dy = 2.5\ kg$$



Nota: O valor médio só existe desde que a soma (caso discreto) ou o integral (caso contínuo) sejam absolutamente convergentes.

Por exemplo, se considerarmos o tempo de uma determinada marca de pilhas referido no exemplo 2.9,

$$E(X) = \int_{100}^{+\infty} x \frac{100}{x^2}\ dx = 100 \int_{100}^{+\infty} \frac{1}{x}\ dx = 100 [\ln x]_{100}^{+\infty} = +\infty$$

Definição 3.2 *Seja X uma v.a. e ϕ uma função real de variável real. Define-se o* ***valor médio*** *ou* ***esperança matemática*** *de $\phi(X)$, por*

a) $E[\phi(X)] = \sum_{i=1}^{\infty} \phi(x_i) P(X = x_i)$, *caso X seja uma v.a. discreta com valores no suporte $D = \{x_1, x_2, \ldots\}$;*

b) $E[\phi(X)] = \int_{-\infty}^{+\infty} \phi(x) f_X(x)\, dx$, *caso X seja uma v.a. contínua com função densidade de probabilidade f_X,*

(desde que existam)

Exemplo 3.3 *Consideremos o exemplo 3.1, e admitamos que o ganho obtido, quando são procurados X aviões por dia, é $\phi(X) = 500\sqrt{X}$ euros. O ganho médio diário será de*

$$\begin{aligned} E\left(500\sqrt{X}\right) &= \sum_{x=0}^{3} 500\sqrt{x} P(X = x) = 500 \times \sqrt{0} \times 0.25 + 500 \times \sqrt{1} \times 0.35 + 500 \times \sqrt{2} \times 0.30 + \\ &+ 500 \times \sqrt{3} \times 0.10 = 473.7345747 \text{ euros.} \end{aligned}$$

Consideremos o exemplo 3.2 e determinemos o valor médio de $\phi(X) = X^2$.

$$E\left(X^2\right) = \int_{-\infty}^{+\infty} x^2 f_X(x)\, dx = \int_2^4 x^2 \frac{1}{2}\, dx = \frac{28}{3}\ kg^2$$

3.1.1 Propriedades do valor médio

Proposição 3.1 *Sejam X e Y v.a.'s, a e b constantes reais e ϕ e φ funções reais de variável real.*

- $E(b) = b$;
- $E(aX + b) = aE(X) + b$;
- $E(X \pm Y) = E(X) \pm E(Y)$;
- $E[\phi(X) \pm \varphi(Y)] = E[\phi(X)] \pm E[\varphi(Y)]$;
- *Se X e Y são v.a.'s independentes, $E(XY) = E(X)E(Y)$.*

Exemplo 3.4 *Relativamente às v.a.'s descritas no exemplo 3.2 e aos valores médios determinados no exemplo 3.3,*

$$E\left(2X + Y - 3X^2 + 10\right) = 2E(X) + E(Y) - 3E\left(X^2\right) + 10 = -9.7$$

Relativamente à primeira parte do exemplo 3.3,

$$E\left(\sqrt{X}\right) = E\left(\frac{1}{500} 500\sqrt{X}\right) = \frac{1}{500} E\left(500\sqrt{X}\right) = 0.947469149$$

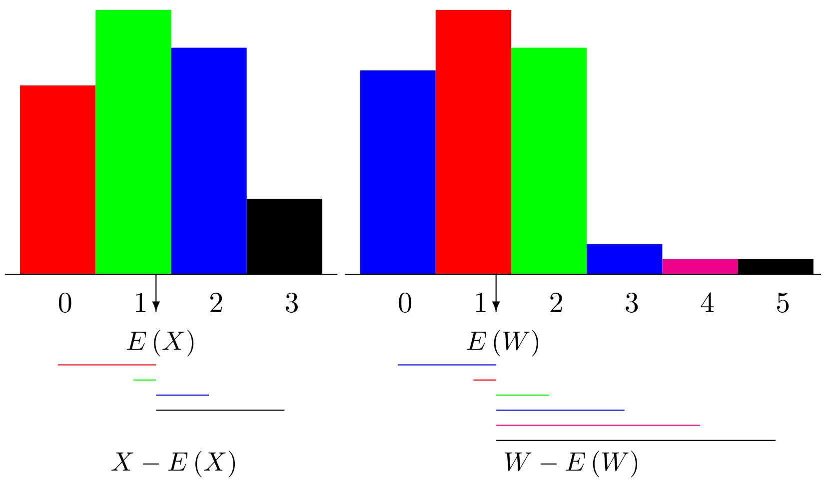
3.2 Variância e desvio padrão

Exemplo 3.5 *Retomemos o exemplo 3.1, e consideremos ainda outra empresa de aluguer de aviões, para a qual a função de probabilidade da procura diária W é,*

$$W\left\{\begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \\ 0.27 & 0.35 & 0.3 & 0.04 & 0.02 & 0.02 \end{array}\right.$$

Se quisermos utilizar o valor médio para analisar a procura diária de aviões nesta e na primeira empresa, isto é, as v.a.'s X e W, constatamos que $E(X) = 1.25$ aviões e $E(W) = 1.25$ aviões.

O valor médio é insuficiente para descrever as características de X e de W. As duas variáveis têm o mesmo ponto de equilíbrio. Contudo, a v.a. W apresenta uma maior dispersão.



Definição 3.3 *Se X é uma v.a., define-se a* ***variância*** *de X por*

$$V(X) = E\left[(X - E(X))^2\right]$$

(desde que exista).

Exemplo 3.6

$$\begin{aligned} V(X) &= E\left[(X-1.25)^2\right] = \sum_{x=0}^{3}(x-1.25)^2 P(X=x) = (0-1.25)^2 \times 0.25 + (1-1.25)^2 \times 0.35 + \\ &+ (2-1.25)^2 \times 0.30 + (3-1.25)^2 \times 0.10 = 0.8875 \ n.^{o2} \ \textit{aviões} \\ V(W) &= E\left[(W-1.25)^2\right] = \sum_{w=0}^{5}(w-1.25)^2 P(W=w) = (0-1.25)^2 \times 0.27 + \ldots + (5-1.25)^2 \times 0.02 = \\ &= 1.1675 \ n.^{o2} \ \textit{aviões} \end{aligned}$$

Proposição 3.2 *Se X é uma v.a., para a qual existe variância, então*

$$V(X) = E\left(X^2\right) - E^2(X)$$

Demonstração: Pela propriedades do valor médio (proposição 3.1),

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

Exemplo 3.7

$$\begin{aligned} E\left(X^2\right) &= \sum_{x=0}^{3} x^2 P(X=x) = 0^2 \times 0.25 + \ldots + 3^2 \times 0.10 = 2.45 \; n.^{o2} \; \textit{aviões} \\ V(X) &= 2.45 - 1.25^2 = 0.8875 \; n.^{o2} \; \textit{aviões} \\ E\left(W^2\right) &= \sum_{w=0}^{5} w^2 P(W=w) = 0^2 \times 0.27 + \ldots + 5^2 \times 0.02 = 2.73 \; n.^{o2} \; \textit{aviões} \\ V(W) &= 2.73 - 1.25^2 = 1.1675 \; n.^{o2} \; \textit{aviões} \end{aligned}$$

Reparou na escala de medição da variância?

Definição 3.4 *Se X é uma v.a., para a qual existe variância, define-se o **desvio padrão** de X por*

$$\sigma(X) = \sqrt{V(X)}$$

Exemplo 3.8

$$\begin{aligned} \sigma(X) &= \sqrt{0.8875} = 0.9420721841 \; n.^{o} \; \textit{de aviões} \\ \sigma(W) &= \sqrt{1.1675} = 1.080509139 \; n.^{o} \; \textit{de aviões} \end{aligned}$$

A variância e o desvio padrão são **medidas de dispersão** (ou **parâmetros de dispersão**) da v.a. a que se aplicam.

Teorema 3.1 *(Desigualdade de Chebychev)*
Se X é uma v.a. para a qual existe variância e $c > 0$ é uma constante real, então

$$P(|X - E(X)| \geq c\sigma(X)) \leq \frac{1}{c^2}$$

A desigualdade de Chebychev também pode ser expressa por

$$P(|X - E(X)| < c\sigma(X)) \geq 1 - \frac{1}{c^2}$$

Nota:

- Para $c = 1.5$, podemos dizer que a probabilidade da v.a. X assumir valores no intervalo $(E(X) - 1.5\sigma(X), E(X) + 1.5\sigma(X))$ é superior a $1 - 1/1.5^2 \approx 0.56$.
- Para $c = 2$, podemos dizer que a probabilidade da v.a. X assumir valores no intervalo $(E(X) - 2\sigma(X), E(X) + 2\sigma(X))$ é superior a $1 - 1/4 = 0.75$.
- Para $c = 3$, podemos dizer que a probabilidade da v.a. X assumir valores no intervalo $(E(X) - 3\sigma(X), E(X) + 3\sigma(X))$ é superior a $1 - 1/9 = 0.89$.

Exemplo 3.9 *Para a v.a. W podemos concluir que, em pelo menos 75% dos dias a procura de aviões se situa no intervalo*

$$\left(1.25 - 2 \times \sqrt{1.1675}, 1.25 + 2 \times \sqrt{1.1675}\right) = (-0.9110, 3.4110) \; \textit{aviões}$$

3.2.1 Propriedades da variância

Proposição 3.3 *Sejam X e Y v.a.'s, a e b constantes reais.*

- $V(b) = 0$;
- $V(aX + b) = a^2 V(X)$;
- *Se X e Y são v.a.'s independentes, $V(X \pm Y) = V(X) + V(Y)$ (Ver proposição 3.5).*

3.3 Covariância e coeficiente de correlação

Exemplo 3.10 *Retomemos o exemplo 2.11 relativo ao par aleatório discreto (X, Y) com*

- X-*o n^o de carros que chegam a um parque de estacionamento, num dado momento*
- Y-*o n^o de lugares vagos neste parque, no mesmo momento*

e admitamos agora que este par aleatório tem um comportamento probabilístico diferente, isto é, tem uma função de probabilidade conjunta

FUNÇÃO DE PROBABILIDADE CONJUNTA DE (X, Y)

$X \setminus Y$	*1*	*2*	*3*	
1	*0.05*	*0.13*	*0.12*	*0.3*
2	*0.1*	*0.06*	*0.14*	*0.3*
3	*0.15*	*0.05*	*0.05*	*0.25*
4	*0.1*	*0.02*	*0.03*	*0.15*
	0.4	*0.26*	*0.34*	*1*

Como se pode constatar, as v.a.'s X *e* Y *não são independentes porque, por exemplo,*

$$P(X = 1; Y = 1) = 0.05 \quad e \quad P(X = 1) \times P(Y = 1) = 0.12$$

Existindo uma relação entre X *e* Y*, podemos tentar medir essa relação com uma nova medida a que se dá o nome de covariância.*

Consideremos o valor esperado de X *e o valor esperado de* Y*,*

$$\begin{aligned} E(X) &= \sum_{x=1}^{4} x\, P(X = x) = 1 \times 0.3 + 2 \times 0.3 + 3 \times 0.25 + 4 \times 0.15 = 2.25 \\ E(Y) &= \sum_{y=1}^{3} y\, P(Y = y) = 1 \times 0.4 + 2 \times 0.26 + 3 \times 0.34 = 1.94 \end{aligned}$$

O quadro da função de probabilidade conjunta de (X, Y) *pode agora ser dividido em 4 zonas,*

$X \setminus Y$	*1*	*2* *3*
1 *2*	A	B
3 *4*	C	D

que se podem interpretar do seguinte modo: se pensarmos que os valores de X *menores que o seu valor esperado* $E(X)$*, são os valores mais pequenos de* X *e que os valores de* X *maiores que o seu valor esperado* $E(X)$*, são os valores maiores de* X*, e pensarmos do mesmo modo para os valores de* Y*, então:*

- *A zona* A *corresponde aos valores menores de* X *que são acompanhados pelos valores menores de* Y*;*
- *A zona* B *corresponde aos valores menores de* X *que são acompanhados pelos valores maiores de* Y*;*
- *A zona* C *corresponde aos valores maiores de* X *que são acompanhados pelos valores menores de* Y*;*
- *A zona* D *corresponde aos valores maiores de* X *que são acompanhados pelos valores maiores de* Y*.*

Então as zonas A e D são as zonas em que as v.a.'s X e Y têm idêntico sentido de crescimento, no sentido em que, quando uma cresce, a outra, com grande probabilidade, também cresce. As zonas B e C são as zonas em que as v.a.'s X e Y têm sentidos de crescimento opostos, no sentido em que, quando uma cresce, a outra, com grande probabilidade, decresce.

Também se constata que:

- *Na zona A,* $X-E(X)$ *tem sinal negativo e* $Y-E(Y)$ *tem sinal negativo, logo* $(X-E(X))(Y-E(Y))$ *tem sinal positivo;*
- *Na zona B,* $X-E(X)$ *tem sinal negativo e* $Y-E(Y)$ *tem sinal positivo, logo* $(X-E(X))(Y-E(Y))$ *tem sinal negativo;*
- *Na zona C,* $X-E(X)$ *tem sinal positivo e* $Y-E(Y)$ *tem sinal negativo, logo* $(X-E(X))(Y-E(Y))$ *tem sinal negativo;*
- *Na zona D,* $X-E(X)$ *tem sinal positivo e* $Y-E(Y)$ *tem sinal positivo, logo* $(X-E(X))(Y-E(Y))$ *tem sinal positivo.*

Em resumo, nas zonas A e D, $(X-E(X))(Y-E(Y))$ *tem sinal positivo e se compararmos com o que anteriormente foi dito sobre estas duas zonas, esse sinal corresponde ao caso em que as v.a.'s X e Y têm idêntico sentido de crescimento.*

Nas zonas B e C, $(X-E(X))(Y-E(Y))$ *tem sinal negativo, correspondendo ao caso em que as v.a.'s X e Y têm sentidos de crescimento opostos.*

Assim, se ponderarmos os valores de $(X-E(X))(Y-E(Y))$ *pelas respectivas probabilidades de ocorrência, ficaremos a saber, em média, qual o sinal preponderante, ou seja, ficaremos a saber, qual o sentido de crescimento preponderante entre as v.a.'s.*

Como tal, define-se a covariância de um par aleatório (X,Y) por

Definição 3.5 *Se* (X,Y) *é um par aleatório, define-se* ***covariância*** *de* (X,Y) *por*

$$cov(X,Y) = E\left[(X-E(X))(Y-E(Y))\right]$$

Neste exemplo,

$$\begin{aligned} cov(X,Y) &= \sum_{x=1}^{4}\sum_{y=1}^{3}(x-2.25)(y-1.94)\ P(X=x;Y=y) = \\ &= (1-2.25)(1-1.94)P(X=1;Y=1)+(2-2.25)(1-1.94)P(X=2;Y=1)+\ldots+ \\ &+ (4-2.25)(3-1.94)P(X=4;Y=3) = -0.295 \end{aligned}$$

Mas a covariância admite ainda outra expressão,

Proposição 3.4 *Seja* (X,Y) *é um par aleatório*

$$cov(X,Y) = E(XY) - E(X)E(Y)$$

que é muito utilizada para efeitos de cálculo.

Exemplo 3.11 *No nosso exemplo, a sua aplicação resulta em*

$$\begin{aligned} E(XY) &= \sum_{x=1}^{4}\sum_{y=1}^{3} xy\,P(X=x;Y=y) = 1\times1\times P(X=1;Y=1)+2\times1\times P(X=2;Y=1) \\ &+ 3\times1\times P(X=3;Y=1)+4\times1\times P(X=4;Y=1)+\ldots+ \\ &+ 4\times3\times P(X=4;Y=3) = 4.07 \\ cov(X,Y) &= E(XY)-E(X)E(Y) = 4.07-2.25\times1.94 = -0.295 \end{aligned}$$

A análise do valor obtido para a covariância permite-nos dizer que, sendo esta negativa, existe uma tendência para que as variáveis tenham sentidos de crescimento opostos, isto é, quando o valor de X aumenta, com grande probabilidade, o valor de Y diminui.

3.3.1 Propriedades da covariância

Proposição 3.5 *Sejam X, Y, W e Z v.a.'s, a,b, c e d constantes reais.*

- $V(X \pm Y) = V(X) + V(Y) \pm 2\, cov(X,Y)$;
- *Se X e Y são v.a.'s independentes, então* $cov(X,Y) = 0$;
- $cov(a + bX, c + dY) = bd\, cov(X,Y)$;
- $cov(aX + bY, cZ + dW) = ac\, cov(X,Z) + ad\, cov(X,W) + bc\, cov(Y,Z) + bd\, cov(Y,W)$.

Contudo, para podermos tecer considerações acerca da "força" da associação entre duas variáveis, não convém usar unicamente a covariância, já que esta medida não é limitada e, como tal, nunca poderemos fazer afirmações do tipo "a covariância é grande" ou "a covariância é pequena". Assim precisamos de uma medida da relação entre as variáveis X e Y, que seja limitada.

Essa medida limitada será o coeficiente de correlação,

Definição 3.6 *Se (X,Y) é um par aleatório, define-se* ***coeficiente de correlação*** *de (X,Y) por*

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)\,V(Y)}}$$

3.3.2 Propriedades do coeficiente de correlação

Proposição 3.6 *Seja (X,Y) um par aleatório,*

- $-1 \leq \rho(X,Y) \leq 1$;
- $|\rho(X,Y)| = 1$ *se, e só se,* $P(Y = a + bX) = 1$, *sendo a e b constantes reais;*
- *Se X e Y são v.a.'s independentes, então* $\rho(X,Y) = 0$.

Nota: O coeficiente de correlação espelha uma relação forte entre duas v.a.'s, desde que atinja valores inferiores a -0.7 ou superiores a 0.7.

Exemplo 3.12 *Voltando ao nosso exemplo,*

$$V(X) = 1.0875 \quad V(Y) = 0.7364 \quad \rho(X,Y) = -0.3296480365$$

e assim concluímos que, existindo relação entre X e Y, ela é muito fraca.

3.4 Outros valores esperados sobre um par aleatório

Definição 3.7 *Seja* (X,Y) *um par aleatório e* $\psi(X,Y)$ *uma função aplicada sobre o par e de valor real. Define-se* ***valor médio*** *ou* ***valor esperado*** *de* $\psi(X,Y)$ *por*

- $E[\psi(X,Y)] = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\psi(x_i,y_j)\,P(X=x_i;Y=y_j)$, *caso o par seja discreto com valores em* $D=\{(x_1,y_1),(x_2,y_2),\ldots\}$;
- $E[\psi(X,Y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\psi(x,y)\,f_{(X,Y)}(x,y)\,dx\,dy$, *caso o par seja contínuo com função densidade de probabilidade conjunta* $f_{(X,Y)}$.

3.5 Outros parâmetros de localização de uma variável aleatória

3.5.1 A mediana

Definição 3.8 *Seja* X *uma v.a. contínua. Designa-se por* ***mediana*** *de* X, *o valor* m_e *de* X *que verifica,*

$$P(X \leq m_e) = 1/2$$

ou, se F *for a função distribuição de* X,

$$F(m_e) = 1/2$$

Exemplo 3.13 *Para a v.a. contínua apresentada no exemplo 2.9, a mediana de* X *terá o valor* $m_e = 200$, *porque*

$$P(X \leq m_e) = 1/2 \Leftrightarrow \int_{-\infty}^{m_e} f_X(x)\,dx = 1/2 \Leftrightarrow \int_{100}^{m_e}\frac{100}{x^2}\,dx = 1/2 \Leftrightarrow$$
$$\Leftrightarrow \quad 1-\frac{100}{m_e} = 1/2 \Leftrightarrow m_e = 200$$

Caso X seja uma v.a. discreta, pode não existir um valor m_e tal que $P(X \leq m_e)$ seja exactamente igual a $1/2$. Daí uma diferente definição de mediana para este tipo de v.a.'s.

Definição 3.9 *Seja* X *uma v.a. discreta. A* ***mediana*** *de* X, *que representamos por* m_e, *é o menor valor de* X *que verifica,*

$$P(X \leq m_e) \geq 1/2$$

Exemplo 3.14 *Para a v.a. discreta* Y, *apresentada no exemplo 3.1, a mediana tem o valor* $m_e = 3$ *porque* $P(Y \leq 2) = 0.45 < 0.5$ *e* $P(Y \leq 3) = 0.75 \geq 0.5$.

3.5.2 O quantil

Definição 3.10 *Seja X uma v.a. contínua. Designa-se por* ***quantil*** *de probabilidade p da v.a. X, o valor χ_p de X que verifica,*

$$P(X \leq \chi_p) = p$$

ou, se F for a função distribuição de X,

$$F(\chi_p) = p$$

Exemplo 3.15 *Para a v.a. contínua apresentada no exemplo 2.9, o quantil de probabilidade 0.9 de X terá o valor $\chi_{0.9} = 1000$, porque*

$$P(X \leq \chi_{0.9}) = 0.9 \Leftrightarrow 1 - \frac{100}{\chi_{0.9}} = 0.9 \Leftrightarrow \chi_{0.9} = 1000$$

Caso X seja uma v.a. discreta, pode não existir um valor χ_p tal que $P(X \leq \chi_p) = p$. Daí uma diferente definição do quantil de probabilidade p para este tipo de v.a.'s.

Definição 3.11 *Seja X uma v.a. discreta. O* ***quantil*** *de probabilidade p de X, que representamos por χ_p, é o menor valor de X que verifica,*

$$P(X \leq \chi_p) \geq p$$

Exemplo 3.16 *Para a v.a. discreta Y, apresentada no exemplo 3.1, o quantil de probabilidade 0.8 tem o valor $\chi_{0.8} = 4$, porque $P(Y \leq 3) = 0.75 < 0.8$ e $P(Y \leq 4) = 0.9 \geq 0.8$.*

3.5.3 A moda

Tal como a designação sugere, a moda é o valor mais "frequente" de uma v.a., ou seja o valor que ocorre com maior probabilidade.

Definição 3.12 *A* ***moda*** *da v.a. X, é o valor m_o tal que*

a) $\max_{x_i \in D} P(X = x_i) = P(X = m_o)$, *caso X seja uma v.a. discreta com valores em $D = \{x_1, x_2, \ldots\}$;*

b) $\max_{x \in \mathbb{R}} f_X(x) = f_X(m_o)$, *caso X seja uma v.a. contínua com função densidade de probabilidade f_X.*

Exemplo 3.17 *Para a v.a. discreta Y do exemplo 3.1, a moda tem o valor $m_o = 2$.*

Para a v.a. contínua descrita no exemplo 2.9 a moda tem o valor $m_o = 100$.

Nota: Caso haja mais do que um valor da v.a. em que ocorra um máximo, dizemos que a moda não existe. (É o que acontece com a v.a. X descrita no exemplo 3.2).

3.6 Outros parâmetros de dispersão de uma variável aleatória

3.6.1 O desvio médio

Definição 3.13 *Seja X uma v.a.. Define-se o* ***desvio médio*** *de X por,*

$$E(|X - E(X)|)$$

Capítulo 4

DISTRIBUIÇÕES IMPORTANTES

4.1 Distribuições discretas

4.1.1 Distribuição Hipergeométrica

Num aquário existem 9 peixes, dos quais 5 estão saudáveis (S) e os restantes 4 estão doentes (D).

Considere-se a seguinte experiência aleatória: extracção ao acaso e **sem reposição** de 3 peixes do aquário e registo do seu estado de saúde.

Associada a esta experiência aleatória, considere-se a seguinte variável aleatória: X - nº de peixes saudáveis na amostra extraída de 3 peixes.

Nota: Repare na natureza dicotómica das características em observação nos peixes, saudável e não saudável.

Pretendemos deduzir a função de probabilidade desta v.a.,

$$X \{$$

Comecemos por relacionar os resultados da experiência com os correspondentes valores de X.

Resultados da experiência	Valores de X
(S, S, S)	3
(S, S, D)	2
(S, D, S)	2
(D, S, S)	2
(S, D, D)	1
(D, S, D)	1
(D, D, S)	1
(D, D, D)	0

Para já podemos completar a função de probabilidade de X com os valores observáveis desta v.a.

$$X \left\{ \begin{array}{cccc} 0 & 1 & 2 & 3 \\ & & & \end{array} \right.$$

Passemos ao cálculo das probabilidades. Por exemplo, se considerarmos o acontecimento $X = 2$, verificamos que é equivalente a ter sido extraída uma das amostras do conjunto, $\{(S, S, D), (S, D, S), (D, S, S)\}$ (de elementos disjuntos dois a dois), pelo que

$$P(X = 2) = P(S, S, D) + P(S, D, S) + P(D, S, S)$$

Ora

$$\left.\begin{array}{l} P(S,S,D)=\frac{5}{9}\times\frac{4}{8}\times\frac{4}{7} \\ P(S,D,S)=\frac{5}{9}\times\frac{4}{8}\times\frac{4}{7} \\ P(D,S,S)=\frac{4}{9}\times\frac{5}{8}\times\frac{4}{7} \end{array}\right\}\Rightarrow P(X=2)=3\times\frac{5}{9}\times\frac{4}{8}\times\frac{4}{7}=\binom{3}{2}\frac{5}{9}\times\frac{4}{8}\times\frac{4}{7}$$

O facto de 3 corresponder a $\binom{3}{2}$, compreender-se-á se pensarmos que 3 é o número de amostras em que se observam 2 peixes saudáveis. Este mesmo número poderá ser determinado pensando que o número de amostras com 2 peixes saudáveis resulta do total de escolhas de 2 posições, de entre as 3 disponíveis, isto a 1ª, a 2ª ou a 3ª no terno que representa a amostra extraída. Assim o total de amostras com dois peixes saudáveis é o total de conjuntos de 2 posições seleccionáveis de entre 3, ou seja $\binom{3}{2}$.

Se repetirmos este processo de cálculo das probabilidades para os restantes valores de X, obtém-se a função de probabilidade

$$X\left\{\begin{array}{cccc} 0 & 1 & 2 & 3 \\ \binom{3}{0}\frac{4}{9}\frac{3}{8}\frac{2}{7} & \binom{3}{1}\frac{5}{9}\frac{4}{8}\frac{3}{7} & \binom{3}{2}\frac{5}{9}\frac{4}{8}\frac{4}{7} & \binom{3}{3}\frac{5}{9}\frac{4}{8}\frac{3}{7} \end{array}\right.$$

Contudo, podemos adoptar outro raciocínio de cálculo das probabilidades. Na verdade, se pensarmos bem, a ordem porque saem os peixes não tem qualquer interesse já que só estamos a contar o nº de peixes saudáveis de entre 3 extraídos. Assim podemos considerar que o resultado da experiência são conjuntos de 3 peixes. Com esta abordagem os diferentes conjuntos que vamos observar são:

Resultados da experiência	Valores de X
$\{S,S,S\}$	3
$\{S,S,D\}$	2
$\{S,D,D\}$	1
$\{D,D,D\}$	0

As probabilidades podem agora ser calculadas usando a lei de Laplace. O número de casos possíveis é o total de conjuntos de 3 peixes que é possível escolher a partir dos 9 que existem no aquário, ou seja $\binom{9}{3}$. Para o acontecimento $\{X=2\}$, o número de casos favoráveis é o total de conjuntos do tipo $\{S,S,D\}$ que podemos formar a partir de 5 peixes saudáveis de um total de 9 peixes, isto é,

$$\binom{5}{2}\binom{9-5}{3-2}$$

Então

$$P(X=2)=\frac{\binom{5}{2}\binom{9-5}{3-2}}{\binom{9}{3}}$$

Concluindo, a função de probabilidade da v.a. X também pode ser escrita por

$$X\left\{\begin{array}{cccc} 0 & 1 & 2 & 3 \\ \frac{\binom{5}{0}\binom{9-5}{3-0}}{\binom{9}{3}} & \frac{\binom{5}{1}\binom{9-5}{3-1}}{\binom{9}{3}} & \frac{\binom{5}{2}\binom{9-5}{3-2}}{\binom{9}{3}} & \frac{\binom{5}{3}\binom{9-5}{3-3}}{\binom{9}{3}} \end{array}\right.$$

Esta variável X diz-se ter distribuição **hipergeométrica de parâmetros** $(\mathbf{9},\mathbf{5},\mathbf{3})$ ou, em escrita abreviada, $\mathbf{X}\sim\mathbf{H}(\mathbf{9},\mathbf{5},\mathbf{3})$.

Características gerais da distribuição hipergeométrica

Numa população **finita** constituída por N elementos, sabemos que M gozam de uma característica A e que os restantes $N-M$ não gozam desta característica A. Considere-se a experiência aleatória que consiste em seleccionar ao acaso e **sem reposição** uma amostra de n elementos da população. Associada a esta experiência aleatória, defina-se a v.a. X-n^{o} de elementos com a característica A, na amostra seleccionada sem reposição. Esta v.a. X tem uma função de probabilidade, que resumidamente, se pode escrever

$$P\left(X=k\right)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \quad \max\left(0,n+M-N\right)\leq k\leq \min\left(n,M\right)$$

e diz-se ter distribuição hipergeométrica de parâmetros (N,M,n) (obviamente $M\leq N$ e $n\leq N$). Dizemos que a distribuição tem parâmetros (N,M,n), porque são as quantidades que é fundamental conhecermos para podermos calcular qualquer probabilidade relativa à v.a. X. A frase "X tem distribuição hipergeométrica de parâmetros (N,M,n)", pode ser escrita abreviadamente $X\sim H\left(N,M,n\right)$.

Coeficientes importantes

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

Observações

- Só podemos seleccionar um número finito de elementos da população, isto é, $n\leq N$.
- O resultado das sucessivas extracções de elementos da população para a amostra não são independentes. Melhor dizendo, a probabilidade de numa extracção sair um elemento com a característica A depende do número de elementos com a característica A que saíram anteriormente.
- Natureza dicotómica do que vamos observar nos elementos extraídos da população, isto é, se têm característica A ou se não têm característica A. Nestes casos, se estamos interessados em observar a presença da característica A, dizemos que, quando é observada, se dá um **sucesso** e que, quando não é observada, se dá um **insucesso**.

4.1.2 Distribuição Binomial

Retomemos o exemplo apresentado na apresentação da distribuição hipergeométrica.

Num aquário existem 9 peixes, dos quais 5 estão saudáveis (S) e os restantes 4 estão doentes (D).

Experiência aleatória: extracção ao acaso e **com reposição** de 3 peixes do aquário e registo do seu estado de saúde.

Consideremos a variável aleatória: X - n^{o} de peixes saudáveis na amostra extraída de 3 peixes.

Pretendemos deduzir a função de probabilidade desta v.a. X.

$$X\left\{\right.$$

Comecemos por relacionar os resultados da experiência com os correspondentes valores de X.

Resultados da experiência	Valores de X
(S,S,S)	3
(S,S,D)	2
(S,D,S)	2
(D,S,S)	2
(S,D,D)	1
(D,S,D)	1
(D,D,S)	1
(D,D,D)	0

Para já podemos completar a função de probabilidade de X com os valores observáveis desta v.a.

$$X\left\{\begin{array}{cccc} 0 & 1 & 2 & 3 \\ & & & \end{array}\right.$$

Passemos ao cálculo das probabilidades. Por exemplo, se considerarmos o acontecimento $X=2$, verificamos que é equivalente a ter sido extraída uma das amostras do conjunto, $\{(S,S,D),(S,D,S),(D,S,S)\}$ (de elementos disjuntos dois a dois), pelo que

$$P(X=2)=P(S,S,D)+P(S,D,S)+P(D,S,S)$$

Ora

$$\left.\begin{array}{l} P(S,S,D)=\frac{5}{9}\times\frac{5}{9}\times\frac{4}{9} \\ P(S,D,S)=\frac{5}{9}\times\frac{4}{9}\times\frac{5}{9} \\ P(D,S,S)=\frac{4}{9}\times\frac{5}{9}\times\frac{5}{9} \end{array}\right\}\Rightarrow P(X=2)=3\times\frac{5}{9}\times\frac{5}{9}\times\frac{4}{9}=\binom{3}{2}\left(\frac{5}{9}\right)^2\left(\frac{4}{9}\right)^{3-2}$$

Se repetirmos este processo de cálculo das probabilidades para os restantes valores de X, obtém-se a função de probabilidade

$$X\left\{\begin{array}{cccc} 0 & 1 & 2 & 3 \\ \binom{3}{0}\frac{4}{9}\frac{4}{9}\frac{4}{9} & \binom{3}{1}\frac{5}{9}\frac{4}{9}\frac{4}{9} & \binom{3}{2}\frac{5}{9}\frac{5}{9}\frac{4}{9} & \binom{3}{3}\frac{5}{9}\frac{5}{9}\frac{5}{9} \end{array}\right.$$

ou ainda

$$X\left\{\begin{array}{cccc} 0 & 1 & 2 & 3 \\ \binom{3}{0}\left(\frac{5}{9}\right)^0\left(\frac{4}{9}\right)^{3-0} & \binom{3}{1}\left(\frac{5}{9}\right)^1\left(\frac{4}{9}\right)^{3-1} & \binom{3}{2}\left(\frac{5}{9}\right)^2\left(\frac{4}{9}\right)^{3-2} & \binom{3}{3}\left(\frac{5}{9}\right)^3\left(\frac{4}{9}\right)^{3-3} \end{array}\right.$$

Esta variável X diz-se ter distribuição **binomial de parâmetros** $\left(\mathbf{3},\frac{\mathbf{5}}{\mathbf{9}}\right)$ ou, abreviadamente, $\mathbf{X}\sim\mathbf{B}\left(\mathbf{3},\frac{\mathbf{5}}{\mathbf{9}}\right)$.

O valor do primeiro parâmetro corresponde ao tamanho da amostra extraída, isto é $n=3$ peixes extraídos e o segundo parâmetro corresponde à probabilidade de, em cada extracção, sair um peixe saudável, isto é $p=\dfrac{5}{9}$.

Evidentemente que considerámos a probabilidade de sair um peixe saudável em cada extracção, porque a nossa v.a. X faz a contagem de peixes saudáveis na amostra de 3 peixes.

Observações:

- Repare que, mesmo sendo finito o n$^{\text{o}}$ de peixes no aquário, como a extracção da amostra é feita com reposição, os peixes disponíveis nunca se esgotam. Podemos então dizer que, para efeitos de extracção, temos um n$^{\text{o}}$ infinito de peixes.
- Também devido ao método de extracção com reposição, mantém-se constante a probabilidade de sair um peixe saudável, em qualquer extracção.
- Também devido ao método de extracção com reposição, os resultados das sucessivas extracções são independentes.

Características gerais da distribuição binomial

Ao realizarmos uma experiência aleatória, estamos apenas interessados em verificar se um determinado acontecimento A se realiza ou não (realização de A ou de $\overline{A}$).

É habitual dizer-se que, quando se observa A, se deu um "*sucesso*" e quando não se realiza A, se deu um "*insucesso*".

Admitamos que é conhecida a probabilidade de realização de A, ou seja é conhecida a probabilidade de se dar um "*sucesso*", que aqui representamos por p,

$$p = P(A) = P(\text{sucesso})$$

Consideremos agora, a repetição n vezes da experiência, nas seguintes condições:

- Mantém-se constante o valor de $p = P(A) = P(\text{sucesso})$ em todas as experiências

Associemos aos resultados das n experiências a v.a.

$$X = \text{n}^{\text{o}} \text{ de sucessos observados nas } n \text{ experiências}$$

ou

$$X = \text{n}^{\text{o}} \text{ de vezes que se observa } A \text{ nas } n \text{ experiências}$$

Esta v.a. X tem uma função de probabilidade,

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n$$

e diz-se ter distribuição **binomial de parâmetros** $(\mathbf{n}, \mathbf{p})$.

A frase "X tem distribuição binomial de parâmetros (n, p)", pode ser escrita de modo abreviado, $\mathbf{X} \sim \mathbf{B}(\mathbf{n}, \mathbf{p})$.

Coeficientes importantes

$$E(X) = np \qquad V(X) = np(1-p)$$

Observações

- Podemos seleccionar um número infinito de elementos da população, quer esta tenha dimensão finita ou infinita;
- O resultado das sucessivas extracções de elementos da população são independentes. Melhor dizendo, a probabilidade de numa extracção sair um elemento com a característica A não depende do número de elementos com a característica A que saíram anteriormente;
- Natureza dicotómica do que vamos observar nos elementos extraídos da população, isto é, se têm característica A ("sucesso") ou se não têm característica A ("insucesso").
- Quando uma experiência consiste em observar se ocorre um "sucesso" ou um "insucesso", designamo-la por prova de Bernoulli.
- Uma sucessão de n provas de Bernoulli é uma experiência aleatória com as seguintes características:
  - Repetição de n *provas de Bernoulli*;
  - As sucessivas *provas de Bernoulli* ocorrem independentemente de prova para prova;
  - A probabilidade de "sucesso" mantém-se constante em todas as n *provas de Bernoulli.*

Teorema 4.1 *Se $X_1, X_2, \ldots, X_k$ são v.a. independentes tais que $X_i \sim B(n_i, p)$, $\quad i = 1, \ldots, k$, então*

$$X_1 + X_2 + \ldots + X_k \sim B(n_1 + n_2 + \ldots + n_k, p)$$

Diferenças fundamentais entre as distribuições hipergeométrica e binomial

Hipergeométrica	Binomial
População finita constituída por N elementos	População infinita
Extracção sem reposição	Extracção com reposição
Sucessivas extracções são não independentes	Sucessivas extracções são independentes

Aproximação da distribuição hipergeométrica pela distribuição binomial

Pensemos agora num lago com $N = 1000$ peixes dos quais $M = 50$ estão saudáveis (S) e os restantes $N - M = 950$ estão doentes (D).

Experiência aleatória: extracção ao acaso e **sem reposição** de $n = 30$ peixes do lago e registo do seu estado de saúde.

Considere-se a v.a.

$$X = \text{n}^{\text{o}} \text{ de peixes saudáveis na amostra extraída de 30 peixes}$$

Evidentemente que a v.a. X tem distribuição hipergeométrica de parâmetros $(1000, 50, 30)$, $X \sim H(1000, 50, 30)$.

Se quisermos calcular a probabilidade de 10 dos peixes seleccionados serem saudáveis, temos

$$\begin{aligned} P(X = 10) &= \frac{\binom{50}{10}\binom{950}{20}}{\binom{1000}{30}} = \\ &= \binom{30}{10} \frac{50}{1000} \frac{49}{999} \frac{48}{998} \frac{47}{997} \frac{46}{996} \frac{45}{995} \frac{44}{994} \frac{43}{993} \frac{42}{992} \frac{41}{991} \times \\ &\times \frac{950}{990} \frac{949}{989} \frac{948}{988} \frac{947}{987} \cdots \frac{932}{972} \frac{931}{971} \end{aligned}$$

O efectivo cálculo desta probabilidade pode constituir um problema porque estão envolvidos números muito elevados. Estamos perante uma situação em que sabemos como calcular a probabilidade mas em que não a podemos calcular com exactidão. Então o melhor será calcularmos um seu valor aproximado, usando um método que permita controlar a qualidade dessa aproximação. Vejamos como o fazer.

Repare que na última expressão de cálculo apresentada,

$$\frac{49}{999} \approx \frac{50}{1000}, \ \frac{48}{998} \approx \frac{50}{1000}, \ldots, \frac{41}{991} \approx \frac{50}{1000}$$

Também

$$\frac{950}{990} \approx \frac{950}{1000}, \ \frac{949}{989} \approx \frac{950}{1000}, \ldots, \frac{931}{971} \approx \frac{950}{1000}$$

Assim

$$\begin{aligned} P(X = 10) &\approx \binom{30}{10} \left(\frac{50}{1000}\right)^{10} \left(\frac{950}{1000}\right)^{20} = \\ &= \binom{30}{10} \left(\frac{50}{1000}\right)^{10} \left(1 - \frac{50}{1000}\right)^{30-10} \end{aligned}$$

Constatamos que a aproximação do valor da probabilidade do acontecimento $X = 10$ corresponde à probabilidade do mesmo acontecimento determinada com a distribuição binomial de parâmetros $n = 30$ e $p = P(\text{peixe saudável}) = \dfrac{50}{1000}$

Conclusão: Podemos aproximar o valor das probabilidades referentes a uma distribuição hipergeométrica pelo valor das probabilidades referentes a uma distribuição binomial (evidentemente, para um mesmo acontecimento $X = k$).

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \approx \binom{n}{k} p^k (1-p)^{n-k},$$

para

$$p = \frac{M}{N} \quad \text{e} \quad \max(0, n + M - N) \leq k \leq \min(n, N)$$

A qualidade da aproximação é "razoável" se o tamanho n da amostra que seleccionámos da nossa população de N elementos, não for muito grande. Como regra, podemos dizer que isto acontece se $n/N \leq 0.1$.

Observações

- Repare que a aproximação resulta de aproximarmos o cálculo da probabilidade de um acontecimento resultante de uma amostragem sem reposição pelo cálculo da probabilidade do mesmo acontecimento como se ele resultasse de uma amostragem com reposição;
- A população que, no caso da distribuição hipergeométrica, é finita, passa, na aproximação à distribuição binomial, a considerar-se infinita;
- Os resultados das sucessivas extracções que, no caso da distribuição hipergeométrica, são acontecimentos não independentes, passam, na aproximação à distribuição binomial, a considerar-se independentes.

4.1.3 Distribuição de Poisson

Exemplo 4.1 *É conhecido que dos indivíduos que têm seguro para determinado tipo de acidente, num ano, 0.0005 morrem deste tipo de acidente.*

Qual a probabilidade de, num ano, a companhia de seguros pagar o seguro deste tipo de acidente a 12 dos 10000 segurados com apólices que cobrem este tipo de acidente.

Defina-se a v.a. $X = n^o$ de segurados que morrem num ano, de entre os 10000.

X é uma v.a. com distribuição $B\,(10000, 0.0005)$ e

$$P\,(X = 12) = \binom{10000}{12} (0.0005)^{12} (1 - 0.0005)^{10000-12} = ??$$

Problema: Sabemos como calcular a probabilidade mas ao fazê-lo efectivamente, podemos ter dificuldades devido ao elevado valor de $\binom{10000}{12}$ e também devido às elevadas potências envolvidas.

Como resolver o problema?

Seja $\{X_n\}$ uma sequência de v.a.'s tais que $X_n \sim B\,(n, p_n)$.

Suponhamos que, à medida que n aumenta, o valor de p_n diminui de modo a que o produto np_n se mantenha estável, isto é constante, com um valor λ. Dito de outro modo, suponhamos que

$$\lim_{n\to\infty} np_n = \lambda > 0$$

Que efeito é que esta condição produzirá se quisermos calcular uma probabilidade para uma v.a. X_n, com n grande?

Se considerarmos $p_n = \lambda/n$,

$$\begin{aligned}
P(X_n = k) &= \binom{n}{k} (p_n)^k (1-p_n)^{n-k} = \\
&= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} = \\
&= \frac{n!}{k!\,(n-k)!} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k} \left(\frac{\lambda}{n}\right)^k = \\
&= \frac{n(n-1)\dots(n-k+1)}{n^k} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k} \frac{\lambda^k}{k!} = \\
&= \frac{1\left(1-\frac{1}{n}\right)\dots\left(1-\frac{k+1}{n}\right)}{\left(1-\frac{\lambda}{n}\right)^k} \left(1-\frac{\lambda}{n}\right)^n \frac{\lambda^k}{k!}
\end{aligned}$$

Ora

$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda},$$

$$\lim_{n\to\infty} \frac{1\left(1-\frac{1}{n}\right)\dots\left(1-\frac{k+1}{n}\right)}{\left(1-\frac{\lambda}{n}\right)^k} = 1$$

e portanto

$$\lim_{n\to\infty} P(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Este resultado permite duas conclusões importantes:

1. Se considerarmos uma v.a. Y que registe o nº de sucessos numa sucessão infinita de experiências e a sua função de probabilidade for

$$P(Y = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

dizemos que Y tem distribuição de **Poisson com parâmetro** λ e, escrevemos de modo abreviado $\mathbf{Y} \sim \mathbf{P}(\lambda)$.

Exercício 4.1 *Verificar que se trata de facto de uma função de probabilidade.*

Coeficientes importantes

$$E(Y) = \lambda \qquad V(Y) = \lambda$$

Aproximação da distribuição binomial pela distribuição de Poisson

2. Se X é uma v.a. com distribuição $B(n,p)$, em que n tem um valor "grande"e p tem um valor "pequeno", então

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}\approx e^{-\lambda}\frac{\lambda^k}{k!},\quad k=0,1,\ldots,n$$

sendo $\lambda=np$.

Isto significa que a distribuição binomial pode ser aproximada pela distribuição de Poisson.

Na prática considera-se "razoável"a aproximação quando $n\geq 30$ e $np\leq 5$. Quando $n\geq 30$ e $np>5$ mas $n(1-p)\leq 5$, aplica-se a aproximação sobre a v.a. $n-X$ que como é evidente tem distribuição $B(n,(1-p))$

Para finalizar, apresentamos dois exemplos.

Exemplo 4.2 *Se aplicarmos esta aproximação ao exemplo 4.1, verificamos que* $n=10000\geq 30$ *e* $np=10000\times 0.0005=5\leq 5$. *Assim, com* $\lambda=10000\times 0.0005=5$,

$$P(X=12)\approx e^{-5}\frac{5^{12}}{12!}=0.00343$$

Suponhamos agora que $X\sim B(40,0.9)$. *Constatamos que* $n=40\geq 30$, $np=36$ *mas* $n(1-p)=4\leq 5$. *Então, com* $\lambda=n(1-p)=4$,

$$P(X=31)=P(40-X=9)\approx e^{-4}\frac{4^9}{9!}=0.013231192$$

Teorema 4.2 *Se* $Y_1,Y_2,\ldots,Y_k$ *são v.a. independentes tais que* $Y_i\sim P(\lambda_i)$, $\quad i=1,\ldots,k$, *então*

$$Y_1+Y_2+\ldots+Y_k\sim P(\lambda_1+\lambda_2+\ldots+\lambda_k)$$

Observações

- Pelo facto da distribuição de Poisson aproximar a distribuição binomial quando $p=P(\text{sucesso})$ é muito pequena, ou seja quando o "sucesso" é um acontecimento raro, a distribuição de Poisson é também designada por distribuição dos acontecimentos raros.

- O parâmetro λ pode ser interpretado como uma taxa de realização de sucessos por unidade. A distribuição do número de sucessos registados em várias unidades ou em fracções da unidade, continua a ser Poisson e a correspondente taxa será determinada como a taxa de sucessos nessas várias unidades ou na fracção da unidade.

  Por exemplo, se durante o período de almoço (das 12 às 14 horas) a chegada de automóveis a um parque se processa a uma taxa de 180 automóveis por hora e tem distribuição de Poisson, então a distribuição do n^o de automóveis que chegam em 15 minutos é Poisson com parâmetro $\lambda=180/4=45$ automóveis. Por sua vez a distribuição do n^o de automóveis que chegam durante o período do almoço é Poisson de parâmetro $\lambda=2\times 180=360$ automóveis.

4.1.4 Processo de Poisson

Consideremos uma variável t (com $t \in \mathbb{R}_0^+$) não aleatória destinada a registar um determinado número de unidades em observação, por exemplo, um período de tempo, uma área, um comprimento, etc.

Seja $N(t)$ uma v.a. que regista o número de sucessos observados num certo período de t unidades.

Se, para cada valor fixo t,

$$P(N(t)=k)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}, \quad k\in\{0,1,2,\ldots,\}, \quad \lambda\in\mathbb{R}^+,$$

isto é, se para cada valor fixo t, $N(t)$ tiver distribuição de Poisson de parâmetro λt, dizemos que $\{N(t)\}_{t\geq 0}$ é um **processo de Poisson de parâmetro** λ.

O parâmetro λ pode ser interpretado como a taxa média de ocorrência de sucessos por cada unidade de observação.

Exemplo 4.3 *Num processo de fabricação de placas de vidro produzem-se pequenas bolhas que se distribuem aleatoriamente pelas placas, com uma densidade média de 0.4 bolhas/m^2. Admitamos que $N(t)$ regista o número de bolhas observadas em placas com t m^2 e que $\{N(t)\}_{t\geq 0}$ é um processo de Poisson de parâmetro $\lambda = 0.4$ bolhas/m^2.*

A probabilidade de, numa placa com 1.5×3.0 $m^2 = 4.5$ m^2, haver pelo menos uma bolha, é

$$P(N(4.5)\geq 1)=1-P(N(4.5)=0)=1-e^{-0.4\times 4.5}\frac{(0.4\times 4.5)^0}{0!}=0.834701$$

Em média, observar-se-ão 2.4 bolhas em placas com $6m^2$, porque

$$E(N(6))=\lambda\times 6=0.4\times 6=2.4.$$

4.1.5 Distribuição Binomial Negativa e Distribuição Geométrica

Consideremos um aquário com 9 peixes, dos quais 5 estão saudáveis (S) e os restantes 4 estão doentes ($D=\bar{S}$) e a seguinte experiência aleatória: extracção ao acaso e **com reposição** de peixes até se conseguir observar 4 peixes saudáveis.

Associemos a esta experiência a v.a. X - nº de peixes extraídos até se observarem 4 peixes saudáveis.

Ao deduzirmos a função de probabilidade desta v.a., começamos por constatar que os seus possíveis valores são todos os inteiros maiores ou iguais a 4. Efectivamente, para se conseguirem observar 4 peixes saudáveis temos de extrair, no mínimo 4 peixes.

Passemos ao cálculo das probabilidades. Para tal, é fundamental perceber que, dado as extracções serem feitas com reposição, o resultados das mesmas constituem acontecimentos independentes e, em cada uma, a probabilidade de se observar um peixe saudável é constante e de valor $P(S)=5/9$.

Consideremos o acontecimento $X=4$.

$$P(X=4)=P(S,S,S,S)=\frac{5}{9}\times\frac{5}{9}\times\frac{5}{9}\times\frac{5}{9}=\left(\frac{5}{9}\right)^4$$

Já para o acontecimento $X = 5$, temos

$$\begin{aligned} P(X=5) &= P(D,S,S,S,S) + P(S,D,S,S,S) + P(S,S,D,S,S) + P(S,S,S,D,S) = \\ &= \frac{4}{9}\times\frac{5}{9}\times\frac{5}{9}\times\frac{5}{9}\times\frac{5}{9}+\frac{5}{9}\times\frac{4}{9}\times\frac{5}{9}\times\frac{5}{9}\times\frac{5}{9}+\frac{5}{9}\times\frac{5}{9}\times\frac{4}{9}\times\frac{5}{9}\times\frac{5}{9}+ \\ &\quad +\frac{5}{9}\times\frac{5}{9}\times\frac{5}{9}\times\frac{4}{9}\times\frac{5}{9}+\frac{5}{9}\times\frac{5}{9}\times\frac{5}{9}\times\frac{4}{9}\times\frac{5}{9} = \\ &= 4\times\left(\frac{5}{9}\right)^4\left(\frac{4}{9}\right)^1 = \binom{5-1}{5-4}\left(\frac{5}{9}\right)^4\left(1-\frac{5}{9}\right)^{5-4} \end{aligned}$$

O factor $\binom{5-1}{5-4}$ explica-se do seguinte modo: Para ser necessário extrair 5 peixes para se conseguirem observar 4 saudáveis, obrigatoriamente na última extracção observa-se um peixe saudável. Restam as 4 primeiras extracções, onde numa delas deverá sair um peixe doente, ou seja deverão sair os restantes $5-4$ peixes que não são saudáveis.

Repetindo o mesmo raciocínio, o acontecimento $X = 6$ terá probabilidade de ocorrência

$$P(X=6) = \binom{6-1}{6-4}\left(\frac{5}{9}\right)^4\left(1-\frac{5}{9}\right)^{6-4}.$$

Generalizando, a função de probabilidade da v.a. X é

$$P(X=k) = \binom{k-1}{k-4}\left(\frac{5}{9}\right)^4\left(1-\frac{5}{9}\right)^{k-4}, \quad k\in\{4,5,6,\ldots\}.$$

e dizemos então que X tem distribuição **binomial negativa com parâmetros** $\left(\mathbf{4},\frac{\mathbf{5}}{\mathbf{9}}\right)$.

Características gerais da distribuição binomial negativa

Ao realizarmos uma experiência aleatória, estamos apenas interessados em verificar se um determinado acontecimento A se realiza ou não (realização de A ou de $\overline{A}$).

É habitual dizer-se que, quando se observa A, se deu um "*sucesso*" e quando não se realiza A, se deu um "*insucesso*".

Admitamos que é conhecida a probabilidade de realização de A, ou seja é conhecida a probabilidade de se dar um "*sucesso*", que aqui representamos por p,

$$p = P(A) = P(\text{sucesso})$$

Consideremos agora, repetições da experiência, nas seguintes condições:

- Mantém-se constante o valor de $p = P(A) = P(\text{sucesso})$ em todas as experiências

Consideremos a v.a.

$$X = \text{n}^{\text{o}} \text{ de experiências a realizar até serem observados } s \text{ sucessos}$$

Esta v.a. X tem uma função de probabilidade,

$$P(X=k) = \binom{k-1}{k-s}p^s(1-p)^{k-s}, \quad k = s, s+1, s+2, \ldots$$

e diz-se ter distribuição **binomial negativa de parâmetros** $(\mathbf{s},\mathbf{p})$.

A frase "X tem distribuição binomial negativa de parâmetros (s,p)", pode ser escrita de modo abreviado, $\mathbf{X}\sim\mathbf{BN}(\mathbf{s},\mathbf{p})$.

Coeficientes importantes

$$E(X) = \frac{s}{p} \qquad V(X) = \frac{s(1-p)}{p^2}$$

Observações

- Podemos seleccionar um número infinito de elementos da população e assim podemos dizer que a população é infinita (nunca se esgota);
- O resultado das sucessivas extracções de elementos da população são independentes. Melhor dizendo, a probabilidade de numa extracção sair um elemento com a característica A não depende do número de elementos com a característica A que saíram anteriormente;
- Natureza dicotómica do que vamos observar nos elementos extraídos da população, isto é, se têm característica A ("sucesso") ou se não têm característica A ("insucesso").
- A probabilidade "sucesso" p é sempre a mesma em cada tentativa.
- Ao inverso da distribuição binomial em que se contam o n^o de sucessos num n^o fixo e conhecido de provas de Bernoulli, na distribuição binomial negativa, o n^o de sucessos é fixo e conhecido e, o que contamos são o n^o de provas de Bernoulli a realizar de modo a se conseguir observar esse n^o de sucessos.

Distribuição Geométrica

A distribuição geométrica é o caso particular da distribuição binomial negativa, quando o número de sucessos a observar é igual a 1. Assim, uma v.a. X com distribuição geométrica conta o número de provas de Bernoulli a realizar até se conseguir observar "sucesso" pela primeira vez. Se $p = P(\text{sucesso})$, dizemos que X tem distribuição geométrica de parâmetro p, e podemos escrever de modo abreviado, $\mathbf{X} \sim \mathbf{G}(\mathbf{p})$.

Função de probabilidade

$$P(= k) = p(1-p)^{k-1}, \; k \in \mathbb{N}.$$

Função de distribuição

$$F_X(x) \equiv P(X \leq x) = \sum_{j=1}^{[x]} P(X = j) = \sum_{j=1}^{[x]} p(1-p)^{j-1} = p\frac{1-(1-p)^{[x]}}{p} = 1-(1-p)^{[x]}, \, x \in \mathbb{R}$$

Coeficientes importantes

$$E(X) = \frac{1}{p} \qquad V(X) = \frac{1-p}{p^2}$$

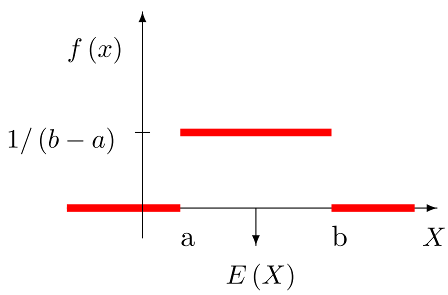
4.2 Distribuições contínuas

4.2.1 Distribuição Uniforme

Esta distribuição utiliza-se quando os valores de certa variável aleatória X, podem ocorrer dentro de um intervalo limitado (aberto, fechado ou semi-aberto) $[a, b]$, e quaisquer dois sub-intervalos com a mesma amplitude têm a mesma probabilidade. Diz-se que X tem distribuição **uniforme no intervalo** $[\mathbf{a}, \mathbf{b}]$ (abreviadamente, escreve-se $\mathbf{X} \sim \mathbf{U}(\mathbf{a}, \mathbf{b})$) e a sua função densidade de probabilidade é

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases}$$

Figura 4.1: Função densidade da distribuição Uniforme



Usando a função densidade de $X \sim U(a, b)$ facilmente se calcula a correspondente função de distribuição, dada por:

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & x \geq b \end{cases}$$

Coeficientes importantes

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

No exemplo 3.2 utilizámos esta distribuição para descrever o peso perdido com o primeiro tipo de dieta. X tinha distribuição uniforme no intervalo $[2, 4]$.

4.2.2 Distribuição Exponencial

O modelo exponencial aplica-se frequentemente quando se pretende estudar os tempos até à ocorrência de avarias, por exemplo, em componentes electrónicas, em que se admite que o tempo que a componente vai durar é independente do tempo que esta já durou, caso a contagem do tempo seja feito a partir do zero. Isto significa que um componente tem um tempo de vida (contado a partir de zero) X com distribuição exponencial, se a sua qualidade se mantém ao longo do tempo, ou seja, se X verifica-se a propriedade:

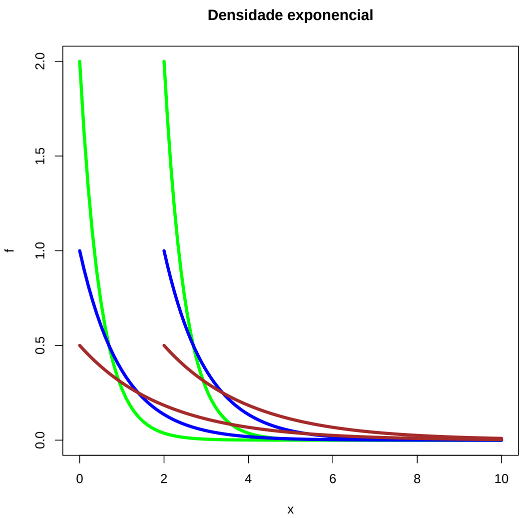
$$P\left(X \geq s+t \,|X \geq s\right) = P\left(X \geq t\right),\ \forall\, s,t \geq 0 \tag{4.2.1}$$

Definição 4.1 *Uma v.a. X tem distribuição* ***exponencial com parâmetros*** (λ, δ)*, se a sua função densidade de probabilidade for*

$$f\left(x\right) = \begin{cases} 0 & x < \lambda \\ \frac{1}{\delta}e^{-(x-\lambda)/\delta} & x \geq \lambda \end{cases}$$

e escreve-se de modo abreviado, $\mathbf{X} \sim \mathbf{E}\left(\lambda, \delta\right)$*. λ e δ deverão ser valores reais e $\delta > 0$.*

Figura 4.2: Exemplos de funções densidade da distribuição exponencial



Observação: A propriedade da falta de memória dada pela expressão (4.2.1), referida acima, apenas é válida quando $\lambda = 0$.

Mais uma vez, podemos deduzir a função distribuição de uma v.a. $X \sim E\left(\lambda, \delta\right)$ usando a correspondente função densidade. Assim temos:

Para $x < \lambda$, $\quad F\left(x\right) = \displaystyle\int_{-\infty}^{x} f\left(x\right) dx = \int_{-\infty}^{x} 0\, dx = 0$

Para $x \geq \lambda$, $\quad F\left(x\right) = \displaystyle\int_{-\infty}^{x} f\left(x\right) dx = \int_{-\infty}^{0} 0\, dx + \int_{\lambda}^{x} \frac{1}{\delta}e^{-(t-\lambda)/\delta}\mathrm{d}t = \left[-e^{-(t-\lambda)/\delta}\right]_{\lambda}^{x} = 1 - e^{-(x-\lambda)/\delta}$

e consequentemente,

$$F\left(x\right) = \begin{cases} 0, & x < \lambda \\ 1 - e^{-(x-\lambda)/\delta}, & x \geq \lambda \end{cases}$$

Coeficientes importantes

$$E(X) = \lambda + \delta \qquad V(X) = \delta^2$$

Exemplo 4.4 *Considere a v.a. X que representa o tempo de espera, em minutos, para ser atendido ao almoço na cantina da FCT/UNL. Admitamos que esta X tem distribuição exponencial e que o tempo médio de espera é de 15 minutos dos quais 1 minuto de espera é sempre "garantido". Qual a probabilidade de, nos cinco dias úteis da semana, em dois deles conseguir ter um tempo de espera superior a 29 minutos?*

Como $\lambda = 1$ minuto e $E(X) = \lambda + \delta = 15$ minutos, então $\delta = 14$ minutos. Repare que o desvio padrão é de 14 minutos.

Calculemos a probabilidade de, num qualquer dia, esperar mais de 29 minutos.

$$p = P(X > 29) = \int_{29}^{+\infty} \frac{1}{14} e^{-(x-1)/14} dx = e^{-2} = 0.135$$

Consideremos agora a v.a. $Y =$n^o de dias com tempo de espera superior a 29 m, de entre 5. Sabemos que Y tem distribuição binomial de parâmetros $(5, 0.135)$. Então a probabilidade pedida tem o valor,

$$P(Y = 2) = \binom{5}{2} (0.135)^2 (1 - 0.135)^{5-2} = 0.117954865$$

Relação entre a distribuição exponencial e o processo de Poisson

Admitamos que $\{N(t)\}_{t \geq 0}$ é um **processo de Poisson de parâmetro** β em que, para valor fixo t, $N(t)$ é uma v.a. que regista o número de sucessos em t unidades de observação.

Para simplicidade de exposição, suponhamos que t representa t períodos de tempo.

Associado a este processo de Poisson, consideremos uma v.a. T que mede o tempo decorrido entre sucessos consecutivos. Evidentemente que T é uma v.a. contínua e os seus valores admissíveis são os reais maiores ou iguais a 0.

Calculemos agora a probabilidade de T assumir valores no intervalo $]t, +\infty], \ t > 0$.

$$\begin{aligned} P(T > t) &= P(T \in]t, +\infty]) = P(\text{não ocorrerem sucessos em } t \text{ períodos de tempo}) = \\ &= P(N(t) = 0) = e^{-\beta t} \frac{(\beta t)^0}{0!} = e^{-\beta t} \end{aligned}$$

Ora, se T é uma v.a. absolutamente contínua com função distribuição dada por

$$F_T(t) = P(T \leq t) = \begin{cases} 0 & t < 0 \\ 1 - e^{-\beta t} & t \geq 0 \end{cases}$$

podemos concluir que a v.a. T tem distribuição **exponencial de parâmetros** $\left(\mathbf{0}, \dfrac{\mathbf{1}}{\beta}\right)$.

Em resumo

Se $\{N(t)\}_{t \geq 0}$ é um **processo de Poisson de parâmetro** β que regista o número de sucessos em t períodos de observação, então a v.a. **T** que mede os períodos decorridos entre sucessos **consecutivos** tem distribuição **exponencial de parâmetros** $(\mathbf{0}, \mathbf{1}/\beta)$.

4.2.3 Distribuição Normal

A distribuição Normal tem grande importância na teoria das probabilidades e na estatística e isto acontece por diversas razões, das quais destacamos:

- Na natureza são inúmeros os fenómenos que são bem descritos por esta distribuição. Por exemplo, a alturas das pessoas de uma grande população, assim como muitos outros tipos de medições como, pesos, volumes, áreas, etc. Também os erros que se cometem ao fazer medições têm frequentemente esta distribuição.

- A distribuição da soma de v.a.'s em grande número e independentes tem uma distribuição aproximadamente normal. Este resultado de grande importância na estatística, será apresentado mais tarde com a designação de Teorema Limite Central.

- As propriedades matemáticas da função densidade de probabilidade desta distribuição, são de tal modo importantes que, muito métodos estatísticos só podem ser utilizados caso se apliquem a fenómenos que têm distribuição normal.

Esta distribuição também é conhecida por distribuição de Gauss, em homenagem ao matemático e físico alemão Carl Gauss (1777-1855) que a apresentou.

Definição 4.2 *Diz-se que a v.a. X tem **distribuição normal de parâmetros** $\left(\mu, \boldsymbol{\sigma^2}\right)$ e escreve-se de modo abreviado, $\mathbf{X} \sim \mathbf{N}\left(\mu, \boldsymbol{\sigma^2}\right)$, se a sua função densidade de probabilidade for,*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R}$$

Os parâmetros μ e σ deverão satisfazer $\mu, \sigma \in \mathbb{R}$ e $\sigma > 0$.

O parâmetro μ indica o ponto de simetria da densidade e o parâmetro σ expressa a dispersão da densidade. Na figura 4.3, o grupo das três curvas à esquerda têm em comum o mesmo valor $\mu = 0$ e o grupo das três curvas à direita têm em comum o mesmo valor $\mu = 24$. As curvas da menos "achatada" à mais "achatada", correspondem a $\sigma = 0.5$, $\sigma = 1$ e a $\sigma = 2$, respectivamente.

Quando $\mu = 0$ e $\sigma = 1$ a distribuição diz-se distribuição **normal reduzida** e normalmente a v.a. associada a esta distribuição é representada por $\mathbf{Z}$, isto é, $\mathbf{Z} \sim \mathbf{N}(\mathbf{0}, \mathbf{1})$. Como veremos adiante, esta distribuição tem um papel muito importante no cálculo de probabilidades de qualquer distribuição normal.

A correspondente função de distribuição é dada pelo integral

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

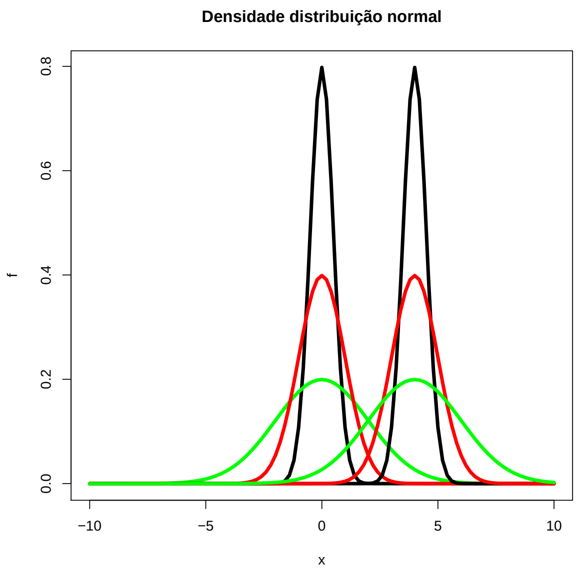
Repare-se que esta função F não tem forma fechada porque a função integranda, ou seja a função densidade f, não é primitivável.

Coeficientes importantes

$$E(X) = \mu \qquad V(X) = \sigma^2$$

Cálculo de probabilidades

Figura 4.3: Exemplos de funções densidade da distribuição Normal



Admitamos que $X \sim N\left(\mu, \sigma^2\right)$ e queremos calcular a $P\left(X \leq b\right)$.

$$P\left(X \leq b\right) = F(b) = \int_{-\infty}^{b} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \, dx$$

Contudo, não sendo conhecida a primitiva da função densidade, não é possível calcular o integral pelos métodos usuais. Terão de ser utilizadas técnicas numéricas de cálculo, que apresentam a desvantagem de serem pesadas e morosas.

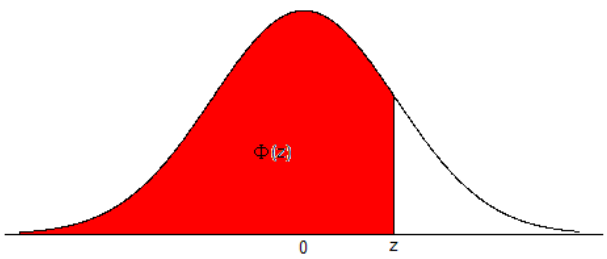
Usando estas técnicas numéricas de cálculo, podemos ter acesso ao valor de probabilidades da distribuição normal reduzida, isto é, para a v.a. $Z \sim N\left(0, 1\right)$. Alguns desses valores encontram-se na dita tabela da função de distribuição normal reduzida.

Nesta tabela encontramos os valores da função

$$\Phi\left(z\right) = P\left(Z \leq z\right) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \, dt, \quad z \in \mathbb{R}$$

designada por **função de distribuição de Z**.

Figura 4.4: Função distribuição de Z



Exemplo 4.5

$$P(Z \leq 1.27) = \Phi(1.27) = 0.8980$$
$$P(Z \geq 0.88) = 1 - P(Z \leq 0.88) = 1 - \Phi(0.88) = 1 - 0.8106 = 0.1894$$

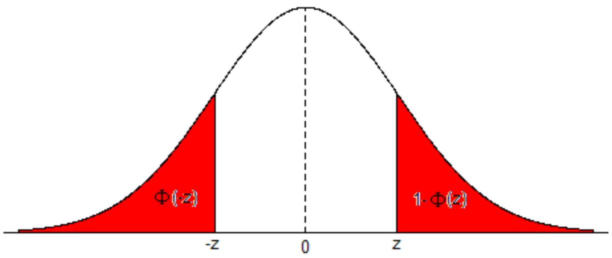
e

$$P(Z \leq -1.27) = ?$$

Seja $Z \sim N(0,1)$ *e* $z \in \mathbb{R}$. *Devido à simetria em torno do ponto 0 da função densidade, podemos deduzir*

$$P(Z \leq -z) = P(X \geq z) \Leftrightarrow P(Z \leq -z) = 1 - P(Z \leq z) \Leftrightarrow \mathbf{\Phi(-z) = 1 - \Phi(z)}$$

Figura 4.5: Consequência da simetria da densidade de Z para a sua função de distribuição



Assim, $P(Z \leq -1.27) = 1 - \Phi(1.27) = 1 - 0.8980 = 0.1020$

Como calcular probabilidades para uma v.a. normal mas não reduzida

Teorema 4.3 *Se* $X \sim N(\mu, \sigma^2)$, *a v.a.* $\dfrac{X-\mu}{\sigma} \sim N(0,1)$.

Exemplo 4.6 *Admitamos que* $X \sim N(1,4)$.

$$\begin{aligned}
P(X \leq 2.98) &= P\left(\frac{X-1}{2} \leq \frac{2.98-1}{2}\right) = P(Z \leq 0.99) = \Phi(0.99) = 0.8389 \\
P(0.40 \leq X < 2.98) &= P\left(\frac{0.40-1}{2} \leq \frac{X-1}{2} < \frac{2.98-1}{2}\right) = P(-0.30 \leq Z < 0.99) = \\
&= P(Z < 0.99) - P(Z \leq -0.30) = \Phi(0.99) - \Phi(-0.30) = \\
&= \Phi(0.99) - (1 - \Phi(0.30)) = 0.8389 - 1 + 0.6179 = 0.4568
\end{aligned}$$

Como calcular quantis de uma v.a. normal

Denotamos por z_p **o quantil de probabilidade** $1-p$ da distribuição normal reduzida, ou seja, se $Z \sim N(0,1)$ e $0 < p < 1$, z_p é o valor de Z que verifica a igualdade

$$P(Z \leq z_p) = 1 - p \quad \text{ou equivalentemente } \Phi(z_p) = 1 - p$$

Exemplo 4.7
$z_{0.2} \approx 0.84$ *porque* $P(Z \leq 0.84) = 0.7995 \approx 0.8$
$z_{0.8} \approx -0.84$ *porque* $P(Z \leq -0.84) = 1 - P(Z \leq 0.84) = 1 - 0.7995 \approx 0.2$

Exemplo 4.8 *Seja $X \sim N(1,4)$ e determinemos o seu quantil de probabilidade 0.8, $\chi_{0.8}$. De acordo com a definição 3.10,*

$$P(X \leq \chi_{0.8}) = 0.8 \quad \Leftrightarrow P\left(\frac{X-1}{2} \leq \frac{\chi_{0.8}-1}{2}\right) = 0.8 \Leftrightarrow P\left(Z \leq \frac{\chi_{0.8}-1}{2}\right) = 0.8 \Leftrightarrow$$
$$\frac{\chi_{0.8}-1}{2} = z_{0.2} = 0.84 \Leftrightarrow \chi_{0.8} = 2 \times 0.84 + 1 = 2.68$$

Outras propriedades da distribuição normal

O próximo teorema generaliza o anterior.

Teorema 4.4 *Se $X \sim N(\mu, \sigma^2)$ e a, b são constantes reais, com $a \neq 0$, então a v.a. $Y = aX + b$ tem distribuição $N(a\mu + b, a^2\sigma^2)$.*

Repare que

$$\begin{aligned} E(Y) &= E(aX+b) = aE(X) + b = a\mu + b \\ V(Y) &= V(aX+b) = a^2V(X) = a^2\sigma^2 \end{aligned}$$

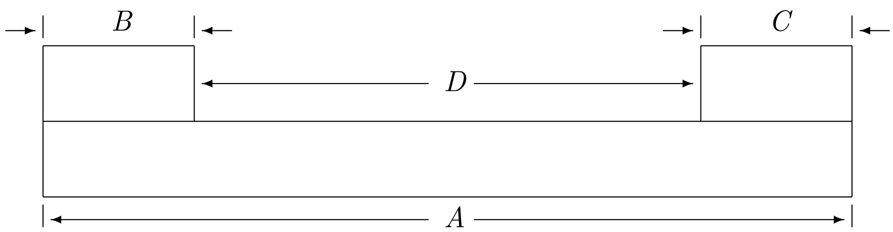
Teorema 4.5 *Sejam $X_1, X_2, \ldots, X_k$ v.a.'s independentes com distribuição $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, k$. Considerem-se $a_1, a_2, \ldots, a_k, b$ constantes reais, com algum $a_i \neq 0$. A v.a.*

$$Y = a_1X_1 + \ldots + a_kX_k + b \sim N\left(a_1\mu_1 + \ldots + a_k\mu_k + b, a_1^2\sigma_1^2 + \ldots + a_k^2\sigma_k^2\right)$$

Repare que

$$\begin{aligned} E(Y) &= E\left(\sum_{i=1}^{k} a_iX_i + b\right) = \sum_{i=1}^{k} a_iE(X_i) + b = \sum_{i=1}^{k} a_i\mu_i + b \\ V(Y) &= V\left(\sum_{i=1}^{k} a_iX_i + b\right) = \sum_{i=1}^{k} a_i^2V(X_i) = \sum_{i=1}^{k} a_i^2\sigma_i^2 \end{aligned}$$

Exemplo 4.9 *Um molde de planificação*



é constituído por três partes cujas larguras A, B e C (em mm) têm as seguintes caraterísticas:

$$A \sim N(10, 0.1) \quad B \sim N(2, 0.05) \quad C \sim N(2, 0.05)$$

e são independentes.

Qual a probabilidade da largura interior do molde, D, ser inferior a 5.9 mm?

Ora $D = A - B - C$ *terá distribuição* $N(E(D), V(D))$. *Como*

$$\begin{aligned} E(D) &= E(A - B - C) = E(A) - E(B) - E(C) = 10 - 2 - 2 = 6 \\ V(D) &= V(A - B - C) = V(A) + V(B) + V(C) = 0.1 + 0.05 + 0.05 = 0.2 \end{aligned}$$

então $D \sim N(6, 0.2)$.

$$P(D < 5.9) = P\left(Z \leq \frac{5.9 - 6}{\sqrt{0.2}}\right) = \Phi(-0.22) = 1 - \Phi(0.22) = 1 - 0.5871 = 0.4129$$

4.2.4 Distribuição Qui-quadrado

Definição 4.3 *Sejam* $Z_1, \ldots, Z_n$ *uma sequência de* n *variáveis aleatórias i.i.d. tais que* $Z_i \sim N(0,1)$. *A variável aleatória*

$$Y = \sum_{i=1}^{n} Z_i^2$$

tem distribuição ***Qui-quadrado*** *com* n *graus de liberdade, e escrevemos* $Y \sim \chi_n^2$.

A função densidade probabilidade de uma variável aleatória $Y \sim \chi_n^2$ é dada por:

$$f(x) = \begin{cases} \frac{1}{\Gamma(n/2)2^{n/2}} e^{-x/2} x^{n/2-1}, & x > 0 \\ 0\ , & x \leq 0 \end{cases}$$

sendo $\Gamma(.)$ a bem conhecida função Gama, definida por:

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

Como exemplo apresentamos o gráfico da função densidade de uma variável aleatória, genérica, com distribuição Qui-quadrado.

Coeficientes importantes

$$E(Y) = n \qquad V(Y) = 2n$$

Notas

1. O cálculo de probabilidades ou de quantis de probabilidade associados a variáveis aleatórias com distribuição Qui-quadrado é feito através de tabelas da função de distribuição, obtidas para diversos graus de liberdade;

2. Denotamos por $\chi^2_{n:\alpha}$ o quantil de probabilidade $1 - \alpha$ da distribuição χ^2_n, ou seja

$$F(\chi^2_{n:\alpha}) = P(Y \leq \chi^2_{n:\alpha}) = 1 - \alpha$$

onde F é a função de distribuição.

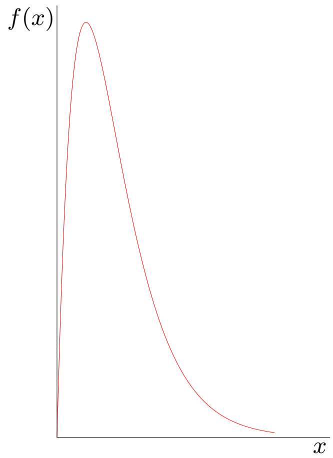


Figura 4.6: Função densidade de $Y \sim \chi^2_n$.

4.2.5 Distribuição t-student

Definição 4.4 *Sejam $Z \sim N(0,1)$ e $Y \sim \chi^2_n$, v.a.'s independentes. A variável aleatória*

$$T = \frac{Z}{\sqrt{Y/n}}$$

diz-se ter uma distribuição ***t-student*** *com n graus de liberdade e escrevemos $T \sim t_n$.*

A sua função densidade probabilidade é dada por:

$$f_n(t) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{n\pi}}(1+t^2/n)^{-(n+1)/2}, \qquad t \in \mathbb{R}$$

Apresenta-se na figura 4.7 o gráfico da função densidade de uma v.a. $T \sim t_n$ genérica.

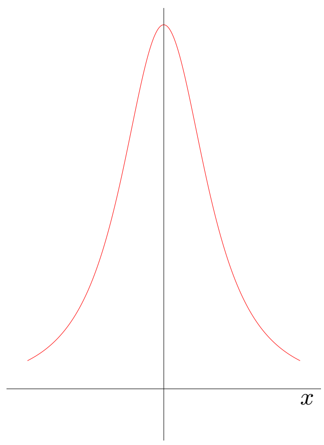


Figura 4.7: Função densidade de $T \sim t_n$.

Coeficientes importantes

$$E\,(T) = 0 \quad , \quad n > 1 \qquad V\,(T) = \frac{n}{n-2} \quad , \quad n > 2$$

Notas

1. À semelhança do que se verifica para a distribuição Qui-quadrado, o cálculo de probabilidades ou de quantis de probabilidade associados a variáveis aleatórias com distribuição t-student também é feito através de tabelas da função de distribuição;

2. Denotamos por $t_{n:\alpha}$ o quantil de probabilidade $1-\alpha$ da distribuição t-student, ou seja

$$F(t_{n:\alpha}) = P(T \leq t_{n:\alpha}) = 1 - \alpha$$

3. A função de densidade da distribuição t-student é simétrica em relação à origem. Assim, $t_{n:\alpha} = -t_{n:1-\alpha}$.

4.2.6 Distribuição F de Fisher

Definição 4.5 *Sejam $C_m \sim \chi^2_m$ e $C_n \sim \chi^2_n$ v.a.'s independentes. A variável aleatória*

$$X = \frac{C_m/m}{C_n/n}$$

*tem uma distribuição **F** (de Fisher ou Snedecor) com m e n graus de liberdade, e escrevemos $X \sim F_{m,n}$.*

A sua função densidade probabilidade é dada por:

$$f_{m,n}(x) = \frac{\Gamma\left(\frac{n+m}{2}\right) m^{m/2} n^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{x^{m/2-1}}{(n+m\,x)^{(m+n)/2}}, \qquad t \in [0, \infty[$$

na figura 4.8 apresentamos um exemplo do gráfico da função densidade de um variável aleatória com distribuição $F_{m,n}$.

Coeficientes importantes

$$E\,(X) = \frac{n}{n-2} \quad , \quad n > 2 \qquad V\,(X) = \frac{2n^2(n+m-2)}{m(n-2)^2(n-4)} \quad , \quad n > 4$$

Notas

1. O cálculo de probabilidades ou de quantis de probabilidade associados a variáveis aleatórias com distribuição F também é feito através de tabelas existentes para o efeito;

2. Denotaremos por $F_{(m,n):\alpha}$ o quantil de probabilidade $1-\alpha$ de uma variável aleatória X com distribuição $F_{m,n}$, ou seja

$$F\left(F_{(m,n):\alpha}\right) = P\left(X \leq F_{(m,n):\alpha}\right) = 1 - \alpha$$

3. Tem-se a seguinte relação entre quantis

$$F_{(m,n):1-\alpha} = \frac{1}{F_{(n,m):\alpha}}$$

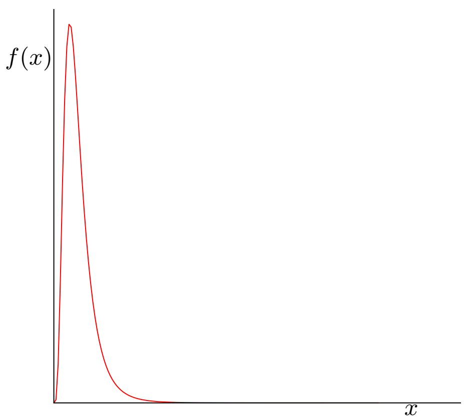


Figura 4.8: Função densidade de $X \sim F_{m,n}$.

Capítulo 5

TEOREMA LIMITE CENTRAL

O Teorema Limite Central tem uma enorme importância na Teoria da Probabilidades e na Estatística porque permite, em condições muito gerais, determinar de modo aproximado, probabilidades relativas à soma ou a médias de variáveis aleatórias. Isto é possível, mesmo que se desconheça a distribuição exacta dessas variáveis aleatórias.

Teorema 5.1 (Teorema Limite Central)

Seja $X_1, X_2, \ldots, X_i, \ldots$ uma sucessão de v.a.'s independentes e identicamente distribuídas (i.i.d.). Seja μ o valor médio comum a todas as v.a.'s e $\sigma^2 \neq 0$ a variância comum a todas as v.a.'s. Considere-se a sucessão das somas parciais $S_n = \sum_{i=1}^{n} X_i,\ n = 1, 2, \ldots$.

Então

$$\lim_{n \to +\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq x\right) = P\left(Z \leq x\right) = \Phi\left(x\right)$$

ou seja a v.a. $\dfrac{S_n - n\mu}{\sqrt{n}\sigma}$ tem uma distribuição que converge para a distribuição normal reduzida, quando $n \to +\infty$. Dito de modo abreviado,

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{L} Z \sim N\left(0,1\right)$$

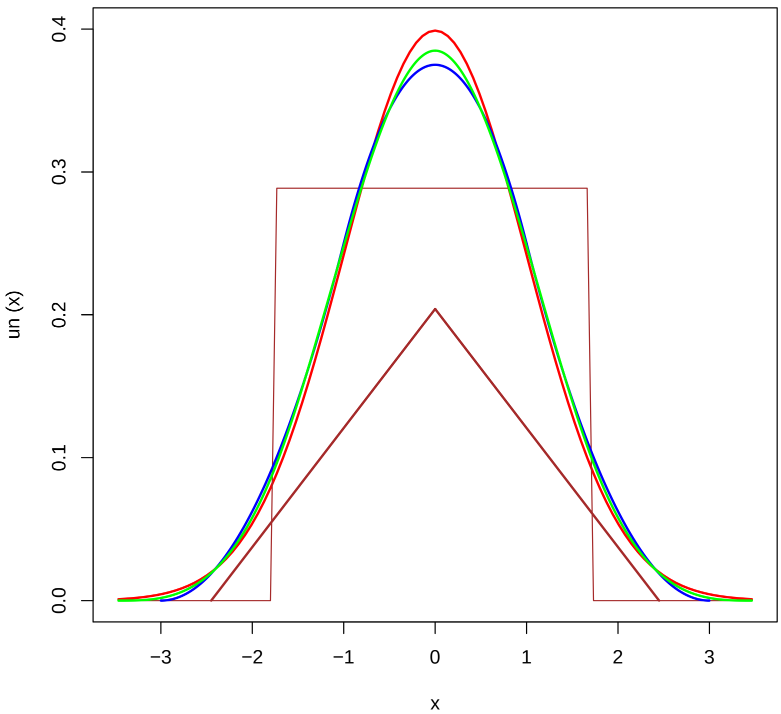
Observações:

- É frequente enunciarmos este teorema dizendo que a v.a. $\dfrac{S_n - n\mu}{\sqrt{n}\sigma}$ tem uma distribuição aproximadamente normal reduzida expressando-o matematicamente por

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \overset{a}{\sim} N\left(0,1\right)$$

- Este teorema tem muito importância em estudos estatísticos e para tal é mais conveniente que possa ser enunciado para uma média aritmética de variáveis aleatórias.

Figura 5.1: Exemplificação do Teorema Limite Central



A bordô, castanho, azul e verde, estão esboçadas as funções densidade da soma (padronizada) de 1, 2, 3 e 4 v.a.'s com distribuição uniforme em $[0,1]$, respectivamente. A vermelho está esboçada a densidade da distribuição $N(0,1)$.

Teorema 5.2 (Teorema Limite Central)

Seja $X_1, X_2, \ldots, X_i, \ldots$ *uma sucessão de v.a.'s independentes e identicamente distribuídas (i.i.d.). Seja* μ *o valor médio comum a todas as v.a.'s e* $\sigma^2 \neq 0$ *a variância comum a todas as v.a.'s. Considere-se a sucessão das médias parciais* $\overline{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i,\ n = 1, 2, \ldots$.

Então

$$\lim_{n \to +\infty} P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq x\right) = P(Z \leq x) = \Phi(x)$$

ou seja a v.a. $\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ *tem uma distribuição que converge para a distribuição normal reduzida, quando* $n \to +\infty$*. Dito de modo abreviado,*

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \overset{L}{\to} Z \sim N(0,1)$$

Neste caso, podemos enunciar o teorema dizendo que a v.a. $\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ tem uma distribuição *aproximadamente* normal reduzida, expressando-o então por

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

Exemplo 5.1 *Num estudo sobre vendas num hipermercado, concluiu-se que a procura diária de arroz (em kg) é uma v.a. com valor médio de 40kg e desvio-padrão de 5kg.*

Tendo sido encomendados 14 500kg de arroz para venda no próximo ano, qual a probabilidade deste stock cobrir a procura de arroz nesse período? (Considere-se um ano com 364 dias).

Sejam X_i *=procura de arroz no dia i, i=1,2,...,364 e admitamos que estas v.a.'s são independentes e identicamente distribuídas.*

Sabemos que:

$$\begin{aligned} E(X_i) &= 40kg, \quad i = 1, 2, \ldots, 364 \\ V(X_i) &= 25kg^2, \quad i = 1, 2, \ldots, 364 \end{aligned}$$

A procura de arroz durante um ano será $S_{364} = \sum_{i=1}^{364} X_i$ *e queremos calcular a* $P(S_{364} \leq 14500)$.

Ignoramos qual a distribuição de S_{364}*, mas como se trata de uma soma de v.a.'s em grande número* $(364 > 30)$*, e sendo satisfeitas as condições do T.L.C., então*

$$\frac{S_{364} - 364 \times 40}{\sqrt{364} \times 5} = \frac{S_{364} - 14560}{\sqrt{364} \times 5} \overset{a}{\sim} N(0, 1)$$

tem uma distribuição bem aproximada pela distribuição normal reduzida. Assim,

$$\begin{aligned} P(S_{364} \leq 14500) &= P\left(\frac{S_{364} - 14560}{\sqrt{364} \times 5} \leq \frac{14500 - 14560}{\sqrt{364} \times 5}\right) \approx \\ &\approx P(Z \leq -0.63) = \Phi(-0.63) = \\ &= 1 - \Phi(0.63) = 1 - 0.7357 = 0.2643 \end{aligned}$$

Conclusão: "É recomendável comprar mais arroz!"

5.1 Aplicações particulares do T.L.C.

5.1.1 Distribuição binomial

Teorema 5.3 *Seja* X *uma v.a. com distribuição binomial de parâmetros* (n, p)*. Se* $n \geq 30$ *e* p *tal que* $np > 5$ *e* $n(1-p) > 5$*, então*

$$\frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{\sqrt{np(1-p)}} \xrightarrow{L} Z \sim N(0, 1)$$

Observação: Como aproximamos a distribuição de uma v.a. discreta pela distribuição de uma v.a. contínua, a aproximação deve ser feita sobre a função distribuição de X, ou seja, sobre probabilidades de acontecimentos do tipo $X \leq k$, sendo k um número inteiro não negativo. Assim sendo,

$$P(X \leq k) = P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{k - np}{\sqrt{np(1-p)}}\right) \approx P\left(Z \leq \frac{k - np}{\sqrt{np(1-p)}}\right), \ k = 0, 1, \ldots$$

Exemplo 5.2 *Suponhamos que* $X \sim B\left(100, \frac{1}{4}\right)$ *e que queremos calcular a* $P(16 \leq X \leq 30)$ *e a* $P(X = 27)$. *Como* $np = \frac{100}{4} = 25 > 5$ *e* $n(1-p) = 100\frac{3}{4} = 75 > 5$,

$$\begin{aligned}
P(16 \leq X \leq 30) &= P(X \leq 30) - P(X < 16) = \\
&= P(X \leq 30) - P(X \leq 15) = \\
&= P\left(\frac{X-25}{\sqrt{18.75}} \leq \frac{30-25}{\sqrt{18.75}}\right) - P\left(\frac{X-25}{\sqrt{18.75}} \leq \frac{15-25}{\sqrt{18.75}}\right) \approx \\
&\approx P(Z \leq 1.15) - P(Z \leq -2.31) = \\
&= \Phi(1.15) - 1 + \Phi(2.31) = 0.8749 - 1 + 0.9896 = 0.8645
\end{aligned}$$

$$\begin{aligned}
P(X = 27) &= P(X \leq 27) - P(X < 27) = \\
&= P(X \leq 27) - P(X \leq 26) = \\
&= P\left(\frac{X-25}{\sqrt{18.75}} \leq \frac{27-25}{\sqrt{18.75}}\right) - P\left(\frac{X-25}{\sqrt{18.75}} \leq \frac{26-25}{\sqrt{18.75}}\right) \approx \\
&\approx P(Z \leq 0.46) - P(Z \leq 0.23) = \\
&= \Phi(0.46) - \Phi(0.23) = 0.6772 - 0.5910 = 0.0862
\end{aligned}$$

Figura 5.2: Ilustração da aproximação da dist. Binomial pela dist. Normal para $n = 2$ e $n = 10$

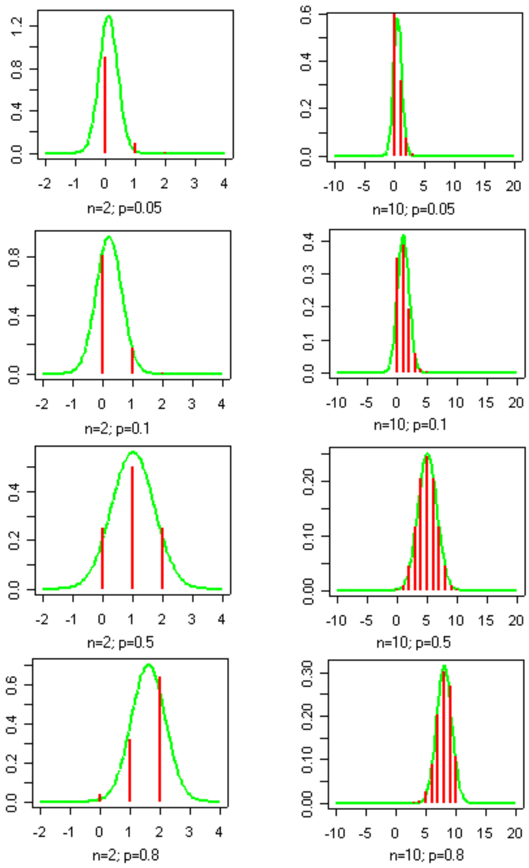


Figura 5.3: Ilustração da aproximação da dist. Binomial pela dist. Normal para $n = 30$

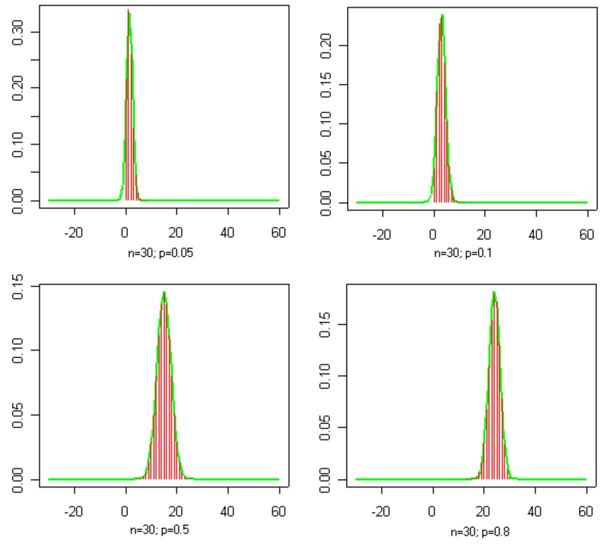
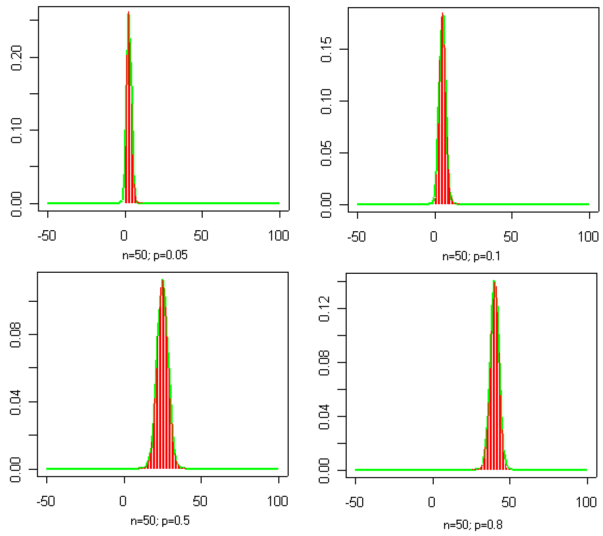


Figura 5.4: Ilustração da aproximação da dist. Binomial pela dist. Normal para $n = 50$



5.1.2 Distribuição de Poisson

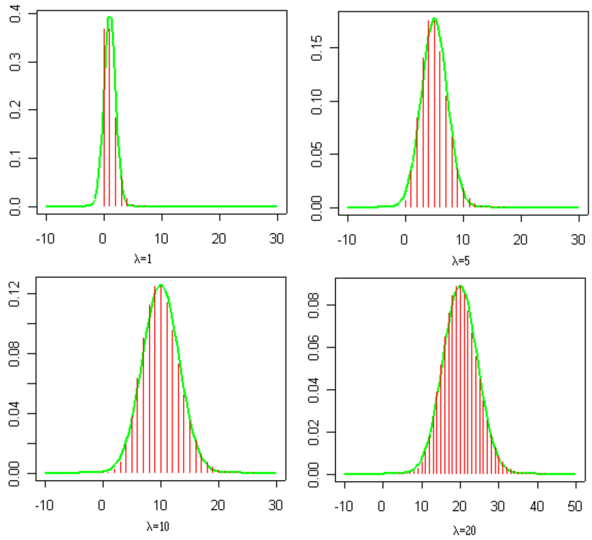
Teorema 5.4 *Seja X uma v.a. com distribuição de Poisson de parâmetro λ. Se $\lambda > 5$,*

$$\frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - \lambda}{\sqrt{\lambda}} \xrightarrow{L} Z \sim N(0, 1)$$

Observação: Como aproximamos a distribuição de uma v.a. discreta pela distribuição de uma v.a. contínua, a aproximação deve ser feita sobre a função distribuição de X, isto é, sobre probabilidades de acontecimentos do tipo $X \leq k$, sendo k um número inteiro não negativo. Assim sendo,

$$P(X \leq k) = P\left(\frac{X - \lambda}{\sqrt{\lambda}} \leq \frac{k - \lambda}{\sqrt{\lambda}}\right) \approx P\left(Z \leq \frac{k - \lambda}{\sqrt{\lambda}}\right), \ k = 0, 1, \ldots$$

Figura 5.5: Ilustração da aproximação da dist. Poisson pela dist. Normal para $\lambda = 1, 5, 10$ e 20



Exemplo 5.3 *Suponha que* $X \sim P(230)$ *e que queremos calcular* $P(X = 241)$.

$$\begin{aligned}
P(X = 241) &= P(X \leq 241) - P(X < 241) = \\
&= P(X \leq 241) - P(X \leq 240) = \\
&= P\left(\frac{X - 230}{\sqrt{230}} \leq \frac{241 - 230}{\sqrt{230}}\right) - P\left(\frac{X - 230}{\sqrt{230}} \leq \frac{240 - 230}{\sqrt{230}}\right) \approx \\
&\approx P(Z \leq 0.73) - P(Z \leq 0.66) = \\
&= \Phi(0.73) - \Phi(0.66) = 0.7673 - 0.7454 = 0.0219
\end{aligned}$$