FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Probabilidades e Estatística
Conteúdos de Probabilidades

2019

Nota introdutória

Este é um documento de apoio ao estudo da componente de Probabilidades para as diversas U.C.'s de Probabilidades e Estatística, servidas pelo Departamento de Matemática da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (DM/FCT/UNL).

O documento tem por base apontamentos elaborados pela Professora Doutora Fátima Miguens nos anos lectivos 2007/2008. Foi revisto pelo Professor Doutor Rui Cardoso nos anos de 2008 a 2010 e objecto das necessárias correções. Entretanto foi actualizado e foi revisto pelo Professores Doutores Ayana Furtado Mateus e Gracinda Rita Guerreiro nos anos de 2011 a 2015. A versão que agora se divulga (2019) apresenta reformulações e novidades nos conteúdos agora expostos. Para já, esta edição foi objecto da revisão por parte do Professor Doutor Luís Ramos e contributos do Professor Doutor Pedro Corte Real. A todos os outros Professores Doutores que contribuíram e venham a contribuir para uma edição final, agradeço a sua colaboração e interesse.

Todos os Professores Doutores mencionados são docentes do DM/FCT/UNL.

A leitura deste texto não dispensa a leitura atenta das obras indicadas como referências bibliográficas nas U.C.'s a que se destina.

Conteúdo

Lista de Figuras

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

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

Nota: 1.1 *A anterior definição de acontecimento está formalmente correta quando o espaço de resultados* Ω *tem um número finito de elementos. Neste caso, se* Ω *tiver* $N \in \mathbb{N}$ *elementos, o total de subconjuntos de* Ω *é* 2^N*.*
Quando Ω *tem um número infinito de elementos, a definição de acontecimento exige uma formulação mais apurada mas que está fora do contexto da disciplina.*
Para efeitos de simplificação do conceito de acontecimento mantemos a definição apresentada.

Definição 1.4 *Denota-se por* $(\Omega, \mathcal{F})$ *o espaço de acontecimentos de uma experiência aleatória cujo universo é* Ω *e* $\mathcal{F}$ *é o conjunto de todos os acontecimentos resultantes da experiência. (ver definição 1.3 e nota 1.1).*

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* $C = \{\textit{Faltou}, \textit{Desistiu}\}$*.*

Definição 1.5 *Diz-se que um acontecimento* $A \in (\Omega, \mathcal{F})$ *ocorre ou se realiza sempre que o resultado da experiência aleatória é um elemento que pertence ao acontecimento* A*.*

1.3 Álgebra de acontecimentos

1. A é *sub-acontecimento* de B se, $A \subseteq B$, isto é, se realização de A implica a realização de B;

2. Dado o acontecimento A, chama-se *acontecimento complementar de* A, e denota-se por $\overline{A}$, o 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. A *Intersecção* dos acontecimentos A e B é um acontecimento que se realiza se, e só se, se realizam em simultâneo os acontecimentos A e B. Representa-se por $A \cap B$;

5. Dados dois acontecimentos A e B, define-se a *diferença A-B* ao acontecimento $A - B = A \cap \overline{B}$, ou seja, à ocorrência de A e não ocorrência de 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 A Probabilidade

Definição 1.6 *(Definição Informal) A* probabilidade *é uma medida (normada) que quantifica a possibilidade (ou a credibilidade, ou a confiança) de ocorrência de um determinado acontecimento.*

Conceito Clássico de Probabilidade

Definição 1.7 (Lei de Laplace) *Se uma experiência aleatória tem $N \in \mathbb{N}$ 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:*

$$P(A) = \frac{N_A}{N} = \frac{n^{o} \text{ de resultados favoráveis à realização do acontecimento } A}{n^{o} \text{ de casos possíveis de observar na experiência}}$$

Exemplo 1.5 *Na continuação do exemplo 1.4*

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

Conceito Frequencista de Probabilidade

E se os resultados da experiência não são igualmente prováveis?

Definição 1.8 *A probabilidade de ocorrência de um acontecimento A é avaliada através do limite (para n repetições infinitas da experiência aleatória), da frequência relativa de observação do acontecimento A nessas repetições. Formalizando, se em n repetições da experiência se observarem n_A ocorrências do acontecimento A, então*

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

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 uma experiência aleatória com o seu espaço de resultados Ω e o espaço de acontecimentos $\mathcal{F}$ associado.

Definição 1.9 *A* probabilidade*, é uma aplicação P que tem por objecto qualquer acontecimento de espaço de acontecimentos $(\Omega, \mathcal{F})$, e como imagem um valor real pertencente ao intervalo $[0,1]$, i.e.,*

$$P: \quad \mathcal{F} \to [0,1]$$

satisfazendo os seguintes axiomas:

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

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

A3. $P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} P(A_i)$, *sendo* $A_1, A_2, \ldots$ *acontecimentos de* $\mathcal{F}$ *e todos disjuntos* *(isto é,* $A_i \cap A_j = \emptyset, \quad i \neq j$*).*

Consequências da axiomática das probabilidades

1. $P\left(\overline{A}\right) = 1 - P(A), \quad \forall A \in \mathcal{F}$;
   Demonstração: $\Omega = A \cup \overline{A}$ e $A \cap \overline{A} = \emptyset$. Pelos axiomas $A2$ e $A3$,
   $P(\Omega) = 1 \Leftrightarrow P(A) + P\left(\overline{A}\right) = 1 \Leftrightarrow P\left(\overline{A}\right) = 1 - P(A)$

2. $P(\emptyset) = 0$; Resultado imediato da consequência 1;

3. $P(A) \leq 1, \forall\, A \in \mathcal{F}$ Resultado imediato da consequência 1

4. $\forall\, A, B \in \mathcal{F}, \quad P(A - B) = P(A) - P(A \cap B)$ ou $P\left(A \cap \overline{B}\right) = P(A) - P(A \cap B)$
   Demonstração: $A = (A \cap B) \cup \left(A \cap \overline{B}\right)$ e $(A \cap B) \cap \left(A \cap \overline{B}\right) = \emptyset$. Pelo axioma $A3$,
   $P(A) = P(A \cap B) + P\left(A \cap \overline{B}\right) \Leftrightarrow P\left(A \cap \overline{B}\right) = P(A) - P(A \cap B)$

5. $P(A \cup B) = P(A) + P(B) - P(A \cap B), \quad \forall\, A, B \in \mathcal{F}$;
   Demonstração: $A \cup B = (A \cap B) \cup \left(A \cap \overline{B}\right) \cup \left(\overline{A} \cap B\right)$ e $A \cap B$, $A \cap \overline{B}$, $\overline{A} \cap B$ são acontecimentos disjuntos.
   Tendo em conta o axioma $A3$ e a consequência 4,
   $P(A \cup B) = P(A \cap B) + P\left(A \cap \overline{B}\right) + P\left(\overline{A} \cap B\right) \Leftrightarrow$
   $= P(A \cap B) + P(A) - P(A \cap B) + P(B) - P(A \cap B) \Leftrightarrow$
   $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

6. $A \subset B \Rightarrow P(A) \leq P(B), \quad \forall\, A, B \in \mathcal{F}$;
   Resultado imediato por aplicação das consequências 3 e 5.

7. $$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i \neq j} P(A_i \cap A_j) + \sum_{i \neq j \neq k} P(A_i \cap A_j \cap A_k) + \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$$
   Pode ser demonstrado por indução em $n \geq 2$.

Definição 1.10 *Denota-se por* $(\Omega, \mathcal{F}, P)$ *o espaço de probabilidade de uma experiência aleatória cujo espaço de acontecimentos é* $(\Omega, \mathcal{F})$ *e no qual se define uma função probabilidade* P*.*

Definição 1.11 *Dois acontecimentos* A *e* B *dizem-se* incompatíveis *se* $P(A \cap B) = 0$*.*

Exemplo 1.6 *Na continuação dos exemplos 1.4 e 1.5*

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

Aplicações

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?

   Considere-se o espaço de resultados constituído por todos os conjuntos de 2 lâmpadas que é possível extraír. O n.º de casos possíveis de observar na experiência é: $\binom{6}{2}$.
   O n.º de casos favoráveis à extração de 2 lâmpadas boas é: $\binom{4}{2}$.

   A probabilidade pedida tem valor: $\frac{\binom{4}{2}}{\binom{6}{2}}$.

   (b) Se retirarmos ao acaso três lâmpadas, qual a probabilidade de todas serem boas?
   Considere-se o espaço de resultados constituído por todos os conjuntos de 3 lâmpadas que é possível extraír. O n.º de casos possíveis de observar na experiência é: $\binom{6}{3}$.
   O n.º de casos favoráveis à extração de 3 lâmpadas boas é: $\binom{4}{3}$.

   A probabilidade pedida tem valor: $\frac{\binom{4}{3}}{\binom{6}{3}}$.

   (c) Retirámos uma lâmpada e verificámos que é boa. Qual a probabilidade de serem boas as duas lâmpadas que a seguir viermos a extrair?

   Considere-se o espaço de resultados constituído por todos os conjuntos de 2 lâmpadas que é possível extraír após a extração da 1.ª.
   O n.º de casos possíveis de observar na experiência é: $\binom{5}{2}$.
   O n.º de casos favoráveis à extração de 2 lâmpadas boas nas 2 últimas extrações é: $\binom{3}{2}$.

   A probabilidade pedida tem valor: $\frac{\binom{3}{2}}{\binom{5}{2}}$.

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. Ambas as equipas vencerão com probabilidade 0.1.

   Considerem-se os acaontecimentos C - a equipa dos Craques vencer o seu desafio, I - a equipa dos Invencíveis vencer o seu desafio.

   Sabemos que: $P(C) = 0.4$, $P(I) = 0.3$ e $P(C \cap I) = 0.1$

   Qual a probabilidade de, no fim da jornada,

   (a) Alguma destas equipas ter ganho?

   $P(C \cup I) = P(C) + P(I) - P(C \cap I) = 0.6$

   (b) Nenhuma delas ter ganho?

   $P\left(\overline{C} \cap \overline{I}\right) = P\left(\overline{C \cup I}\right) = 1 - P(C \cup I) = 0.4$

   (c) Somente os Invencíveis ganharem o respectivo desafio?

   $P\left(I \cap \overline{C}\right) = P(I) - P(I \cap C) = 0.2$

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.

   Após a escolha ao acaso de um aluno, considerem-se os acontecimentos fulcrais:
   F - se praticante de futebol; B - ser praticante de basquetebol; A - ser praticante de andebol.

   Sabemos que: $P(F) = 0.3 \quad P(B) = 0.2 \quad P(A) = 0.2$
   $P(F \cap B) = 0.05 \quad P(F \cap A) = 0.1 \quad P(B \cap A) = 0.05 \quad P(F \cap B \cap A) = 0.02$

   A probabilidade do aluno escolhido ser

   (a) Praticante de futebol ou de andebol é:
   $P(F \cup A) = P(F) + P(A) - P(F \cap A) = 0.4$

   (b) Um atleta é:
   $P(F \cup B \cup A) = P(F) + P(B) + P(A) - P(F \cap B) - P(F \cap A) - P(B \cap A) + P(F \cap B \cap A) = 0.52$

   (c) Apenas jogador de futebol é:
   $P\left(F \cap \overline{B} \cap \overline{A}\right) = P\left(F \cap \overline{B \cup A}\right) = P(F) - P[F \cap (B \cup A)]$
   $= P(F) - P[(F \cap B) \cup (F \cap A)] = P(F) - [P(F \cap B) + P(F \cap A) - P(F \cap B \cap A)] = 0.17$

1.5 Probabilidade condicional

São conhecidos os seguintes valores resultantes da classificação por género e por situação de empregabilidade de 200 indivíduos (maiores de idade):

	Empregado	Desempregado	Total
Feminino	94	10	104
Masculino	88	8	96
Total	182	18	200

- Se seleccionarmos um indivíduo ao acaso, de entre os 200, qual é a probabilidade de estar desempregado?

  Considere o acontecimento D - o indivíduo estar desempregado

  $P(D) = \dfrac{18}{200}$

- Se seleccionarmos ao acaso um indivíduo do género masculino, qual é a probabilidade de estar desempregado?

  Considere o acontecimento M - o indivíduo ser do género masculino

  $$P\left(D\left|M\right.\right)=\frac{8}{96}$$

  Por sua vez,

  $$P\left(D\left|M\right.\right)=\frac{8}{96}=\frac{\frac{8}{200}}{\frac{96}{200}}=\frac{P\left(D\cap M\right)}{P\left(M\right)}$$

Na escolha aleatória de um indivíduo do género masculino, o universo deixa de ser Ω - respeitante à escolha ao acaso de um indivíduo de entre os 200, e passa agora a ser o universo M - respeitante aos indivíduos do género masculino embalagem.

Este novo acontecimento é denotado por

$$D\left|M\right.$$

em que a notação $\cdot\left|M\right.$ serve para informar que o espaço de resultados foi alterado, isto é, deixa de ser $\Omega=\{\text{Todos os 200 indivíduos}\}$ e passa agora a ser $M=\{\text{Todos os indivíduos do género masculino}\}$.

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

$$\frac{P\left(D\cap M\right)}{P\left(M\right)}$$

Adotando a notação proposta para o novo acontecimento,

$$P\left(D\left|M\right.\right)=\frac{P\left(D\cap M\right)}{P\left(M\right)}$$

Em todas as definições, teoremas e exemplos que se seguem até ao final deste capítulo, consideram-se acontecimentos pertencentes a um espaço de acontecimentos $(\Omega,\mathcal{F})$.

Definição 1.12 *Sejam A e B acontecimentos de um espaço de probabilidade $(\Omega,\mathcal{F},P)$ garantindo-se que $P\left(B\right)>0$. 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*

$$P\left(A\left|B\right.\right)=\frac{P\left(A\cap B\right)}{P\left(B\right)}$$

Exemplo 1.7 *Na continuação da aplicação 3 da secção anterior:*
Se escolhermos um aluno atleta, qual a probabilidade de ser:

- *Um jogador de futebol ou de andebol?*

  $$P\left(F\cup A\left|F\cup B\cup A\right.\right)=\frac{P\left[\left(F\cup A\right)\cap\left(F\cup B\cup A\right)\right]}{P\left(F\cup B\cup A\right)}=\frac{P\left(F\cup A\right)}{P\left(F\cup B\cup A\right)}=\frac{0.4}{0.52}\approx 0.7692$$

- *Apenas jogador de futebol?*

$$P\left(F \cap \overline{B} \cap \overline{A} \,|F \cup B \cup A\right) = \frac{P\left(F \cap \overline{B} \cap \overline{A}\right)}{P\left(F \cup B \cup A\right)} = \frac{0.17}{0.52} = 0.3269$$

Exercício 1.1 *Provar que se B é um acontecimento do espaço de probabilidade $(\Omega, \mathcal{F}, P)$ tal que $P(B) > 0$, então $P(\cdot\,|B)$ é uma função de probabilidade.*

Teorema 1.1 (Teorema da probabilidade composta) *Se A, B são acontecimentos de $(\Omega, \mathcal{F}, P)$, tais que $P(A) > 0$ e $P(B) > 0$, então*

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

Este teorema é extensível a mais acontecimentos de $(\Omega, \mathcal{F}, P)$. O exercício seguinte ilustra uma dessas extensões.

Exercício 1.2 *Sejam A, B e C acontecimentos de um espaço de probailidade $(\Omega, \mathcal{F}, P)$ tais que $P(A) > 0$ e $P(A \cap B) > 0$. Mostre que $P(A \cap B \cap C) = P(C\,|A \cap B)\,P(B\,|A)\,P(A)$*

1.5.1 Teorema da probabilidade total; Teorema de Bayes

Definição 1.13 *Os acontecimentos $\{B_1, B_2, \ldots, B_n\}$ constituem uma* ***partição*** *do espaço de acontecimentos $(\Omega, \mathcal{F})$ 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 acontecimentos $(\Omega, \mathcal{F})$, com $P(B_i) > 0$, para $i = 1, 2, \ldots, n$, dado um qualquer acontecimento $A \in (\Omega, \mathcal{F})$, 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 acontecimentos $(\Omega, \mathcal{F})$, com $P(B_i) > 0$, para $i = 1, 2, \ldots, n$, dado um qualquer acontecimento $A \in (\Omega, \mathcal{F})$, 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 1.8 *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".*
Se escolhermos ao acaso um lote do fornecimento total, sabemos que:

$P(B_1) = 0.45,\ P(B_2) = 0.25,\ P(B_3) = 0.30$
$P(D | B_1) = 0.05,\ P(D | B_2) = 0.07,\ P(D | B_3) = 0.10.$

- *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?*

  $$P(B_1 | D) = \frac{P(D | B_1) P(B_1)}{P(D)} = 0.32$$

  $$P(B_2 | D) = \frac{P(D | B_2) P(B_2)}{P(D)} = 0.25$$

  $$P(B_3 | D) = 1 - P(B_1 | D) - P(B_2 | D) = 0.43$$

  Devemos reclamar a origem B_3 por ser a mais provável.

1.6 Independência de acontecimentos

Definição 1.14 *Sejam A e B acontecimentos de um espaço de probabilidade $(\Omega, \mathcal{F}, P)$, garantindo-se que $P(B) > 0$.*
Se o conhecimento de que o acontecimento 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,

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

Definição 1.15 *Os acontecimentos A e B, de um espaço de probabilidade $(\Omega, \mathcal{F}, P)$, dizem-se* ***independentes*** *se, e só se*

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

Teorema 1.4 *Se A e B são acontecimentos de $(\Omega, \mathcal{F}, P)$ e* ***independentes****, também o são A e $\overline{B}$, $\overline{A}$ e B e ainda $\overline{A}$ e $\overline{B}$.*

Definição 1.16 *Sejam $A_1, A_2, \ldots A_r$ acontecimentos de $(\Omega, \mathcal{F}, P)$. $A_1, A_2, \ldots A_r$ são acontecimentos independentes, se e só se, para qualquer conjunto de acontecimentos $A_{k_1}, A_{k_2}, \ldots$*

$$P\left(A_{k_1} \cap A_{k_2} \cap A_{k_p}\right) = P(A_{k_1}) \times P(A_{k_2}) \times \ldots \times P(A_{k_r})$$

Exemplo 1.9 *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?*

Solução: Represente-se por A o acontecimento acertar no alvo em qualquer tiro. $P(A) = 0.6$.

Tendo em conta a independência do acerto em qualquer tiro:

a) $P\left(\underbrace{\overline{A}\cap\cdots\cap\overline{A}}_{9\ ocorr\hat{e}ncias\ de\overline{A}}\cap A\right)=\underbrace{P\left(\overline{A}\right)\times\cdots\times P\left(\overline{A}\right)}_{9factores}\times P\left(A\right)=P\left(A\right)P\left(\overline{A}\right)^{9}=0.000157286$

b) $P\left(A\cap\overline{A}\cap\overline{A}\right)+P\left(\overline{A}\cap A\cap\overline{A}\right)+P\left(\overline{A}\cap\overline{A}\cap A\right)=3\times P\left(A\right)P\left(\overline{A}\right)^{2}=0.288$

c) $P\left(A\cap A\cap\overline{A}\cap\overline{A}\cap A\right)+P\left(A\cap\overline{A}\cap A\cap\overline{A}\cap A\right)+\cdots+P\left(\overline{A}\cap\overline{A}\cap A\cap A\cap A\right)=\binom{5}{3}P\left(A\right)^{3}P\left(\overline{A}\right)^{2}=0.3456$

d) *Comecemos por determinar a probabilidade de serem necessários menos de 4 tiros para se acertarem dois.*

$$\underbrace{P\left(A\cap A\right)}_{Em\ 2\ tiros\ acertar\ 2}+\underbrace{P\left(\overline{A}\cap A\cap A\right)+P\left(A\cap\overline{A}\cap A\right)}_{Em\ 3\ tiros\ acertar\ 2}=P\left(A\right)^{2}+\binom{2}{1}P\left(A\right)^{2}P\left(\overline{A}\right)=0.648$$

Resposta: $1-0.648=0.352$

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 o espaço de acontecimentos $\mathcal{F}$ de um universo Ω, se associarmos a cada acontecimento $A \in \mathcal{F}$ um conjunto de números reais $X(A)$, estamos a definir uma função $X : \mathcal{F} \to \mathbb{R}$.

Sendo $A \in \mathcal{F}$ um acontecimento, chamamos *imagem de A por X*, ao conjunto de valores reais que X assume para os elementos ω de A, isto é,

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

Por outro lado, queremos que a um subconjunto de números reais $E \subset \mathbb{R}$, seja possível atribuir uma probabilidade. Para tal, teremos que assegurar que o subconjunto $X^{-1}(E)$ formado por todos os elementos $\omega \in \Omega$ tais que $X(\omega) \in E$,

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

é um acontecimento de $(\Omega, \mathcal{F})$. $X^{-1}(E)$ 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 $(\Omega, \mathcal{F})$ o espaço de acontecimentos de uma experiência aleatória. Uma* ***variável aleatória*** *X é uma função $X : \mathcal{F} \longrightarrow \mathbb{R}$ tal que, para qualquer intervalo real I, $E = X^{-1}(I)$ é um acontecimento de $(\Omega, \mathcal{F})$ e por isso*

$$P(X \in I) = P(E)$$

existe e pode ser determinada.

A definição que se segue é mais rigorosa e fundamenta-se em conhecimentos mais profundas sobre Análise Matemática.

Definição 2.2 *Seja $(\Omega, \mathcal{F})$ o espaço de acontecimentos de uma experiência aleatória. Uma* ***variável aleatória*** *X é uma função $X : \mathcal{F} \longrightarrow \mathbb{R}$ tal que $A_x = \{\omega \in \Omega : \ X(\omega) \leq x\}$, com $x \in \mathbb{R}$, é um acontecimento de $(\Omega, \mathcal{F})$.*

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: $X = n^{\underline{o}}$ *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 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 muito útil e importante, no cálculo de probabilidades para uma variável aleatória.

Definição 2.3 *A função real de variável real,* F_X, *definida por*

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

designa-se por ***função distribuição*** *da variável aleatória* X.

Nota 1 A função de distribuição tem domínio $\mathbb{R}$ e $F_X(x) = P(X \in \,]-\infty, x]), \ \forall\, x \in \mathbb{R}$.

Nota 2 Esta função acumula valores de probabilidade pois, dados dois reais x e y, com $x > y$, então $]-\infty, x] = \,]-\infty, y] \cup \,]y, x]$, e por isso

$$F_X(x) = P(X \leq x) = P(X \leq y) + P(x < X \leq y) = F_X(y) + \underbrace{P(x < X \leq y)}_{\geq 0} \tag{2.2.1}$$

A função distribuição goza das seguintes propriedades:

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

Proposição 2.3 *Qualquer função distribuição F_X satisfaz as seguintes propriedades:*

I. *é uma função não decrescente em $\mathbb{R}$;*

II. *é uma função contínua à direita em $\mathbb{R}$;*

III. $F_X(-\infty) = \lim\limits_{x \to -\infty} F_X(x) = 0$ *e* $F_X(+\infty) = \lim\limits_{x \to +\infty} F_X(x) = 1$

Nota 3

O conhecimento da função de distribuição permite-nos determinar qualquer probabilidade relativa aos valores de uma v.a. X.

A. Para qualquer real x, podemos usar a função distribuição para determinarmos a $P(X < x)$. Para tal, e denotando por $F_X(x^-)$, o limite à esquerda de F_X no real x,

$$P(X < x) = F_X\left(x^-\right) \tag{2.2.2}$$

A demonstração passa por se considerar um real positivo h e pelas seguintes deduções

$$P(X < x) = \lim_{h \to 0} P(X \leq x - h) = \lim_{h \to 0} F_X(x - h) = F_X\left(x^-\right)$$

B. Para qualquer real a,

$$P(X = a) = F_X(a) - F_X\left(a^-\right) \tag{2.2.3}$$

porque $P(X = a) = P(X \leq a) - P(X < a) = F_X(a) - F_X(a^-)$

C. Dados os reais a e b, com $a < b$,

C1 $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$

C2 $P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F_X(b) - F_X(a^-)$

C3 $P(a \leq X < b) = P(X < b) - P(X < a) = F_X(b^-) - F_X(a^-)$

C4 $P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F_X(b) - F_X(a^-)$

2.3 Variável aleatória discreta

Definição 2.4 *Seja X uma variável aleatória e o conjunto*

$$S_X = \{a \in \mathbb{R} : P(X = a) > 0\}.$$

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

Definição 2.5 *Seja X uma v.a. discreta. Chama-se* ***função de probabilidade*** *(f.p.) de X à função definida no suporte $S_X = \{a_1, a_2, \ldots\}$ dos valores de X e pelas respectivas probabilidades. Uma representação usual para a função de probabilidade da v.a. X, é:*

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

Propriedades da função de probabilidade

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

Cálculo de probabilidades

Dado um qualquer intervalo real, I,

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

2.3.1 Cálculo da função distribuição

A função distribuição é determinada por

$$F_X(x) = P(X \leq x) = \sum_{a_i \in S_X \cap \{a_i \leq x\}} P(X = a_i)$$

Exemplo 2.5 *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 é:*

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

e

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

A função distribuição da v.a. X é:

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

porque

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

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

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

Recorrendo apenas a esta função distribuição:

- $P(X \leq 2) = F_X(2) = 26/27$
- $P(X < 2) = F_X(2^-) = 20/27$
- $P(0.7 \leq X \leq 2.6) = P(X \leq 2.6) - P(X < 0.7) = F_X(2.6) - F_X(0.7^-) = \frac{26}{27} - \frac{8}{27} = \frac{19}{27}$
- $P(1 < X \leq 2) = P(X \leq 2) - P(X \leq 1) = F_X(2) - F_X(1) = \frac{26}{27} - \frac{20}{27} = \frac{6}{27}$

Notas importantes

Como podemos constatar pelo exemplo anterior, a função distribuição de uma variável discreta, possui, para além das propriedades gerais já enunciadas na secção 2.2, outras características que importa realçar:

- trata-se de uma função em "escada";
- os valores reais onde ocorrem discontinuidades ("saltos") são os valores observáveis da variável aleatória, isto é, os valores do suporte de X;
- a amplitude do "salto" num ponto de descontinuidade a corresponde à probabilidade $P(X = a)$ porque

$$P(X = a) = P(X \leq a) - P(X < a) = F_X(a) - F_X(a^-)$$

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

Exemplo 2.6 *Seja X uma variável aleatória com função distribuição*

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

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 e por isso o suporte de X é $S_X = \{-0.6, 2, 5.7\}$. A amplitude dos saltos correspondem à probabilidade de observação destes valores, ou seja

$$\begin{aligned} P(X = -0.6) &= F_X(-0.6) - F_X\left(0.6^-\right) = 0.2 - 0 = 0.2 \\ P(X = 2) &= F_X(2) - F_X\left(2^-\right) = 0.8 - 0.2 = 0.6 \\ P(X = 5.7) &= F_X(5.7) - F_X\left(5.7^-\right) = 1 - 0.8 = 0.2 \end{aligned}$$

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

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

2.4 Variável aleatória absolutamente contínua

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

$$P(X = a) = 0, \quad \forall\, a \in \mathbb{R} \tag{2.4.4}$$

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

$$P(X \in I) = \int_I f_X(x)\, dx \tag{2.4.5}$$

Ao conjunto de valores reais $S_X = \{x \in \mathbb{R} : f_X(x) > 0\}$ 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

Se X é v.a. absolutamente contínua:

Nota 1 Dado um qualquer intervalo real, I,

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

Como se trata do integral de uma função não negativa e o integral é sempre absolutamente convergente, então a $P\left(X \in I\right)$, corresponde ao valor da área da função f_X no intervalo I.

Nota 2 Sendo X uma v.a. absolutamente contínua, e dada a sua caracterização,

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

porque

$$P\left(X \leq a\right) = P\left(X < a\right) + \underbrace{P\left(X = a\right)}_{=0} = P\left(X < a\right), \quad \forall\, a \in \mathbb{R}$$

Nota 3 Dados dois reais a e b, com $a < b$,

$$P\left(a \leq X \leq b\right) = P\left(a < X \leq b\right) = P\left(a \leq X < b\right) = P\left(a < X < b\right)$$

Nota 4 Se X é v.a. absolutamente contínua, e para $h \in \mathbb{R}^+$

$$\lim_{h \to 0} \frac{P\left(a < X \leq a + h\right)}{h} \quad = \quad \lim_{h \to 0} \frac{\int_a^{a+h} f_X\left(x\right) dx}{h} = f\left(a\right).$$

Assim,

$$P\left(a < X \leq a + h\right) \approx h\, f_X\left(a\right), \quad h \in \mathbb{R}^+$$

Nota 5 Se X é v.a. absolutamente contínua com função densidade f_X e função distribuição F_X,

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

2.4.1 Cálculo da função distribuição

A função distribuição é determinada por

$$F_X\left(x\right) = P\left(X \leq x\right) = \int_{-\infty}^{x} f_X\left(t\right)\ dt, \quad \forall\, x \in \mathbb{R} \tag{2.4.6}$$

Nota 6 Se X é v.a. absolutamente contínua, a respectiva função distribuição é uma **função contínua** em $\mathbb{R}$.

A demonstração passa por usarmos as propriedades gerais de uma função distribuição apresentadas na secção 2.2 e pela caracterização 2.4.4 de uma v.a. absolutamente contínua.

Para qualquer real a,

$$P(X \leq a) = P(X < a) + \underbrace{P(X = a)}_{=0} \Leftrightarrow P(X \leq a) = P(X < a) \Leftrightarrow F_X(a) = F_X(a^-).$$

Isto é, a função distribuição é contínua à esquerda em $\mathbb{R}$.

Por sua vez, como qualquer função distribuição é contínua à direita em $\mathbb{R}$, então a função distribuição de uma v.a. absolutamente contínua é uma função contínua em $\mathbb{R}$.

Exemplo 2.7 *Suponhamos que o tempo de vida (em horas) de uma determinada marca de pilhas, X, é uma v.a. com função densidade de probabilidade:*

$$f_X(x) = \begin{cases} c/x^2 & x \in]100, +\infty[\\ 0 & x \in]-\infty, 100] \end{cases}, \quad c \in \mathbb{R} \setminus \{0\}$$

a) c pode ter um valor qualquer?

1. *Como se exige que* $f_X(x) \geq 0, \quad \forall x \in \mathbb{R} \Rightarrow c \geq 0$
2. *Como se exige que* $\int_{-\infty}^{+\infty} f_X(x)\, dx = 1$, *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 probabilidade de uma pilha durar mais de 500 horas?

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

c) Qual a probabilidade de uma pilha durar mais de 120 horas mas não menos de 110 horas?

$$P(110 < X < 120) = \int_{110}^{120} \frac{100}{x^2} dx = 100\left[-\frac{1}{x}\right]_{110}^{120} = 5/66$$

d) Determinação da função distribuição, F_X

- *Se* $x \leq 100, \quad F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = \int_{-\infty}^{x} 0\, dt = 0$
- *Se* $x > 100, \quad F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = \int_{-\infty}^{100} 0\, dt + \int_{100}^{x} \frac{100}{t^2}\, dt = 100\left[-\frac{1}{t}\right]_{100}^{x} = 1 - \frac{100}{x}$

$$F_X(x) = \begin{cases} 0, & x \leq 100 \\ 1 - \frac{100}{x}, & x > 100 \end{cases}$$

e) *Determinação das probabilidades referidas nas alíneas b) e c), recorrendo exclusivamente à função distribuição:*

$P\left(X > 500\right) = 1 - P\left(X \leq 500\right) = 1 - F_X\left(500\right) = 1 - \left(1 - \frac{100}{500}\right) = 0.2$

$P(110 < X < 120) = P\left(X < 120\right) - P\left(X \leq 110\right) = F_X\left(120^-\right) - F_X\left(110\right) =$

$= F_X\left(120\right) - F_X\left(110\right) = \left(1 - \frac{100}{120}\right) - \left(1 - \frac{100}{110}\right) = 5/66$

Capítulo 3

PAR ALEATÓRIO

Exemplo 3.1 *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 $\mathbb{R}^p$ (ou um seu subconjunto).

Definição 3.1 *Um vector* $\boldsymbol{X} \equiv (X_1, X_2, \ldots, X_p)$ *diz-se um* ***vector aleatório*** *de dimensão p se $X_1, X_2, \ldots, X_p$ são v.a.'s aleatórias definidas num mesmo espaço de acontecimentos* $(\Omega, \mathcal{F})$*.*

Quando pretendemos estudar apenas duas v.a.'s, 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.

3.1 Par aleatório discreto

Definição 3.2 *Seja (X, Y) um par aleatório discreto. Chama-se* ***função de probabilidade conjunta*** *(f.p.c) de (X, Y) à função definida no suporte*

$$S_{(X,Y)} = \{(x_1, y_1), (x_2, y_2), \ldots\}$$

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. $0 \leq p_{ij} \leq 1, \quad \forall (x_i, y_j) \in S_{(X,Y)}$
2. $\sum\sum_{(x_i,y_j)\in S_{(X,Y)}} p_{ij} = 1$

Exemplo 3.2 *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 DO P.A. (X, Y)

$X \setminus Y$	1	2	3
1	0.05	0.03	0.02
2	0.1	0.06	0.04
3	0.15	0.09	0.06
4	0.2	0.12	0.08

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

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

FUNÇÃO DE PROBABILIDADE CONJUNTA E MARGINAIS DO P.A. (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*	
	$P(Y=1)$	$P(Y=2)$	$P(Y=3)$		

Definição 3.3 *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*, por, respectivamente:*

$$\begin{aligned} p_{i\cdot} &= P(X=x_i)=\sum_{y_j\in S_{(X,Y)}} P(X=x_i;Y=y_j)=\sum_{y_j\in S_{(X,Y)}} p_{ij}, \quad \forall\, x_i\in S_{(X,Y)} \\ p_{\cdot j} &= P(Y=y_j)=\sum_{x_i\in S_{(X,Y)}} P(X=x_i;Y=y_j)=\sum_{x_i\in S_{(X,Y)}} p_{ij}, \quad \forall\, y_j\in S_{(X,Y)} \end{aligned}$$

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

Exemplo 3.3 *Na continuação do exemplo 3.2*

FUNÇÃO DE PROBABILIDADE MARGINAL DA V.A. 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 DA V.A. Y

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

3.1.1 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 estes 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 3.4 *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,*

$$P(X = x_i; Y = y_j) = P(X = x_i)\, P(Y = y_j)\,, \quad \forall (x_i, y_j) \in S_{(X,Y)}$$

ou numa notação simplificada, se, e só se,

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

Exemplo 3.4 *Na continuação do exemplo 3.2*

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

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 nº de carros que chegam ao parque de estacionamento, num dado momento, não tem qualquer relação probabilística com o nº de lugares vagos, no mesmo momento.

3.2 Par aleatório contínuo

Definição 3.5 *Um par aleatório (X,Y) diz-se* ***absolutamente contínuo*** *(ou* ***contínuo****) se o conjunto*

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

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****), devendo satisfazer as seguintes condições:*

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

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 sendo $f_{(X,Y)}$ 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 3.6 *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, por, respectivamente:*

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

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

3.2.1 Independência das variáveis de um par aleatório contínuo

Definição 3.7 *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 3.5 *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*

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

a) Qual o valor de c?

$$f_{(X,Y)}(x,y) \geq 0, \quad \forall\,(x,y) \in \mathbb{R}^2 \quad \Rightarrow \quad c \geq 0$$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{(X,Y)}(x,y)\,dxdy = 1 \quad \Leftrightarrow \quad \int_0^3\int_0^2 cx^2y\,dxdy = 1 \Leftrightarrow c = 1/18$$

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

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

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?

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

Conclusão: As v.a.'s X e Y são independentes.

e) Com a informação da alínea anterior, a alínea b) poder-se-ia determinar do seguinte 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 4

MOMENTOS E PARÂMETROS

4.1 Valor médio ou valor esperado

Exemplo 4.1 *Uma empresa de aluguer de aviões sabe que a procura diária X tem carácter aleatório e estima a sua função de probabilidade por*

$$X\left\{\begin{array}{cccc} 0 & 1 & 2 & 3 \\ 0.25 & 0.35 & 0.30 & 0.10 \end{array}\right.$$

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 **valor esperado** de X e que se representa por $E(X)$.*

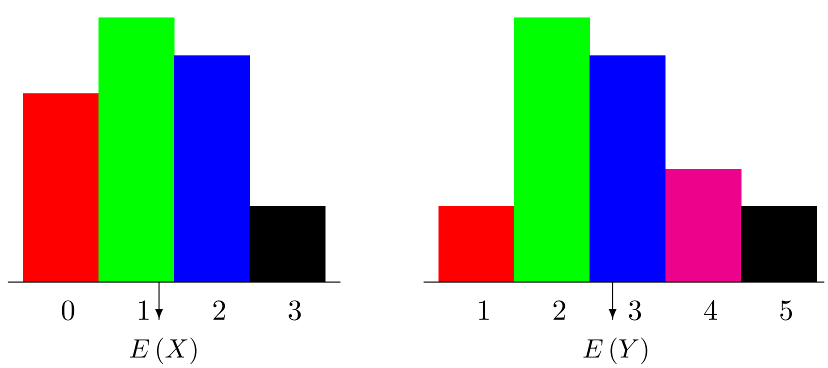
Devido à segunda designação, a quantidade que acabámos de calcular também pode ser enunciada por: "O número esperado de aviões procurados diariamente é de 1.25.

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

$$Y\left\{\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \\ 0.10 & 0.35 & 0.30 & 0.15 & 0.10 \end{array}\right.$$

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 4.1 *Se X é uma v.a., define-se o* ***valor médio*** *(ou* ***valor esperado*** *ou ainda* ***esperança matemática****) de X, por*

Se X é v.a. discreta: $E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i)$,

sendo $S_X = \{x_1, x_2, \ldots\}$ o suporte de X e garantindo-se que
$E(|X|) = \sum_{i=1}^{\infty} |x_i| P(X = x_i) < +\infty$;

Se X é v.a. absolutamente contínua: $E(X) = \int_{-\infty}^{+\infty} x f_X(x)\ dx$,

sendo f_X a função densidade de probabilidade e garantindo-se que

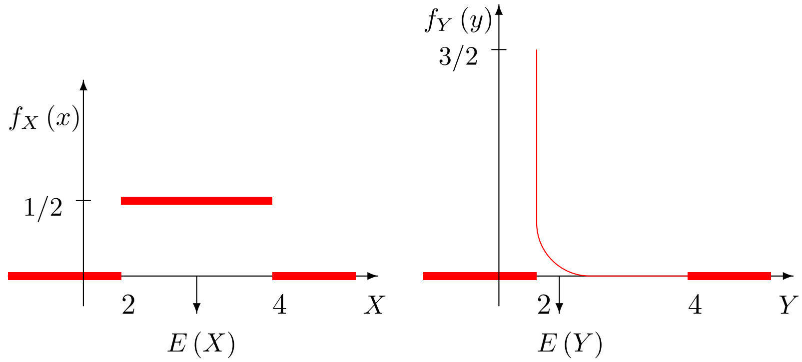
$$E(|X|) = \int_{-\infty}^{+\infty} |x|\, f_X(x)\ dx < +\infty.$$

Exemplo 4.2 *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 uniforme repartição de probabilidade, entre 2 e 4 kg. Quantos quilos espera perder nas duas semanas?*

$$\begin{aligned} f_X(x) &= \begin{cases} 1/2 & 2 \le x \le 4 \\ 0 & o.v.\ de\ x \end{cases} \\ E(X) &= \int_{-\infty}^{+\infty} x f_X(x)\ dx = \int_2^4 x \frac{1}{2}\ dx = 3\ kg \end{aligned}$$

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

$$\begin{aligned} f_Y(y) &= \begin{cases} \frac{3}{8}(y-4)^2 & 2 \le y \le 4 \\ 0 & o.v.\ de\ y \end{cases} \\ E(Y) &= \int_{-\infty}^{+\infty} y f_Y(y)\ dy = \int_2^4 y \frac{3}{8}(y-4)^2\ dy = 2.5\ kg \end{aligned}$$



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 duração da marca de pilhas referido no exemplo 2.7,

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

e por isso não existe valor médio da v.a. X.

Definição 4.2 *Seja X uma v.a. e ϕ uma função real de variável real. Define-se o* ***valor médio*** *(****valor esperado*** *ou* ***esperança matemática****) de $\phi(X)$, por*

a) Caso X seja uma v.a. discreta com valores no suporte $S_X = \{x_1, x_2, \ldots\}$

$$E[\phi(X)] = \sum_{i=1}^{\infty}\phi(x_i)\,P(X = x_i);$$

b) Caso X seja uma v.a. absolutamente contínua com função densidade de probabilidade f_X

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

(desde que existam)

Exemplo 4.3 *Consideremos o exemplo 4.1, 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 4.2 e determinemos o valor médio de $\phi(X) = X^2$.

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

Definição 4.3 *Seja (X,Y) um p.a. e ϕ uma função de $\mathbb{R}^2$ em $\mathbb{R}$. Define-se o* ***valor médio*** *(****valor esperado*** *ou* ***esperança matemática****) de $\phi(X,Y)$, por*

a) Caso (X,Y) seja um p.a. discreto com valores no suporte $S_{(X,Y)} = \{(x_i, y_j)\}, \quad i,j = 1,2,\ldots$

$$E[\phi(X,Y)] = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\phi(x_i,y_j)\,P(X = x_i; Y = y_j);$$

b) Caso (X,Y) seja um p.a. absolutamente contínuo com função densidade de probabilidade conjunta $f_{(X,Y)}$

$$E[\phi(X,y)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\phi(x,y)\,f_{(X,Y)}(x,y)\;dx\,dy,$$

(desde que existam)

4.1.1 Propriedades do valor médio

Elencamos propriedades importantes relativas ao operador valor médio. As demonstrações das mais importantes serão feitas dentro do contexto apropriado.

Proposição 4.1 *Sejam X e Y v.a.'s, a e b constantes reais e g, h são funções reais de variável real.*

- $E(b) = b$;
- $E(aX + b) = aE(X) + b$;
- $E(g(X) + h(X)] = E[g(X)] + E[h(X)]$;
- $E[g(X) + h(Y)] = E[g(X)] + E[h(Y)]$;
- *Se X e Y são v.a.'s independentes, então $E(XY) = E(X)E(Y)$. (ver proposição 4.5)*

Exemplo 4.4 *Relativamente às v.a.'s descritas no exemplo 4.2 e aos valores médios determinados no exemplo 4.3,*

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

e

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

Quanto ao par aleatório descrito no exemplo 3.2, tem-se:

- $E(X) = \sum_{x=1}^{4} x \times P(X = x) = 3$ *n.º carros a chegar/momento*
- $E(1/X) = \sum_{x=1}^{4} \frac{1}{x} \times P(X = x) = 0.4$
- $E(Y) = \sum_{y=1}^{3} y \times P(Y = y) = 1.7$ *n.º lugares estacionamento/momento*
- $E(XY) = \sum_{x=1}^{4} \sum_{y=1}^{3} xy \times P(X = x; Y = y) = 5.1$
- $E(X - Y) = \sum_{x=1}^{4} \sum_{y=1}^{3} (x - y) \times P(X = x; Y = y) = E(X) - E(Y) = 1.3$
- $E\left(\frac{Y}{X}\right) = \sum_{x=1}^{4} \sum_{y=1}^{3} \frac{y}{x} \times P(X = x; Y = y) = 0.68$

Terminamos este exemplo, chamando a atenção para que, no exemplo 3.4 foi comprovada a independência do p.a. (X, Y). Por isso, e só por isso, poderíamos obter, de modo alternativo:

- $E(XY) = E(X)E(Y) = 3 \times 1.7 = 5.1$
- $E(Y/X) = E(Y)E(1/X) = 1.7 \times 0.4 = 0.68$

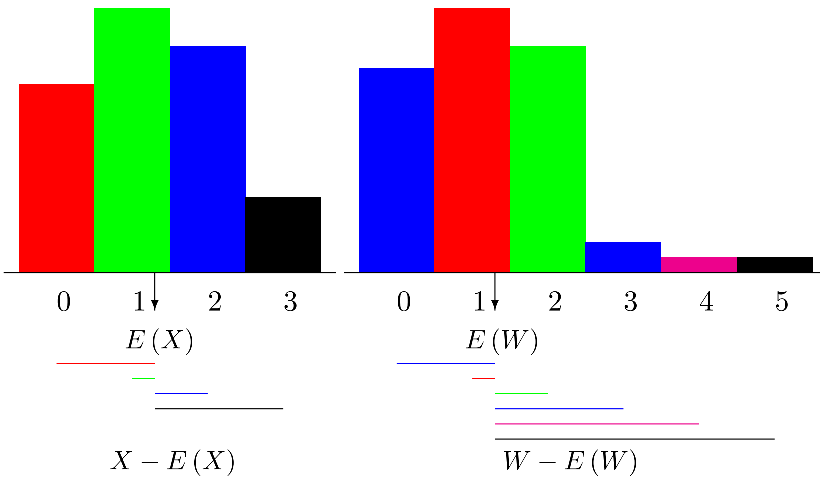
4.2 Variância e desvio padrão

Exemplo 4.5 *Retomemos o exemplo 4.1, 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 quantidades $E(X)$ e $E(W)$, constatamos que $E(X) = 1.25$ aviões e $E(W) = 1.25$ aviões.

O valor médio, só por si, é insuficiente para discriminar as características das v.a.'s X e de W. As duas variáveis têm o mesmo ponto de equilíbrio. Contudo, a v.a. W apresenta uma maior dispersão.



Definição 4.4 *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 4.6

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

Proposição 4.2 *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 4.1),

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

Exemplo 4.7

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

Reparou na escala de medição da variância?

A variância afere a dispersão de uma v.a.. Contudo, a escala de medição do seu valor é sempre o quadrado da escala de medição da v.a..

Para efeitos de avaliação da dispersão de uma v.a., na mesma escala de medição da mesma v.a. propõe-se o desvio padrão.

Definição 4.5 *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 4.8

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

Conclusão: Tendo em conta apenas estas duas quantidades, poderemos concluir que a procura diária por aviões é mais dispersa (imprevisível) na segunda empresa descrita.

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

4.2.1 Propriedades da variância

Proposição 4.3 *Sejam X e Y v.a.'s, a e b constantes reais.*

- $V(b) = 0$;
- $V(aX + b) = a^2 V(X)$;
- $V(X \pm Y) = V(X) + V(Y) \pm 2\, cov(X, y)$

Teorema 4.1 *(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 $c = 1$, podemos garantir que os valores da v.a. X pertencem ao intervalo

  $]E(X) - \sigma(X), E(X) - \sigma(X)[$,

  porque $P(|X - E(X)| < 1 \times \sigma(X)) \geq 1 - \frac{1}{1^2} = 0$
- 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 4.9 *Para a v.a. W podemos concluir que, em pelo menos 75% dos dias a procura de aviões se situa no intervalo*

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

4.3 Covariância e coeficiente de correlação

Exemplo 4.10 *Retomemos o exemplo 3.2 relativo ao par aleatório discreto* (X, Y) *sendo*

- 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*

e admitamos agora que este par aleatório tem um comportamento probabilístico diferente, isto é, tem a função de probabilidade conjunta e as funções de probabilidade marginais que se apresentam na tabela seguinte:

FUNÇÕES DE PROBABILIDADE DO P.A. (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*,*

$$E(X) = \sum_{x=1}^{4} x\, P(X = x) = 1 \times 0.3 + 2 \times 0.3 + 3 \times 0.25 + 4 \times 0.15 = 2.25$$

$$E(Y) = \sum_{y=1}^{3} y\, P(Y = y) = 1 \times 0.4 + 2 \times 0.26 + 3 \times 0.34 = 1.94$$

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

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

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

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

Exemplo 4.11 *Continuando o exemplo 4.10,*

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

Mas a covariância admite ainda outra expressão, que é muito utilizada para efeitos de cálculo.

Proposição 4.4 *Seja* (X, Y) *um par aleatório*

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

Demonstração: Considerem-se as seguintes identificações: $a \equiv E\,(X)$ e $b \equiv E\,(Y)$.

Pela definição 4.6 de covariância,

$$cov\,(X, Y) = E\,[(X - a)\,(Y - b)] = E\,(XY - a\,Y - b\,X + ab) =$$

$$= E\,(XY) - aE\,(Y) - bE\,(X) + ab = E\,(XY) - ab - ab + ab = E\,(XY) - ab$$

Exemplo 4.12 *Na continuação do exemplo 4.10, a sua aplicação resulta em*

$$\begin{aligned}
E\,(XY) &= \sum_{x=1}^{4}\sum_{y=1}^{3} xy\,P\,(X = x; Y = y) = 1 \times 1 \times P\,(X = 1; Y = 1) + 2 \times 1 \times P\,(X = 2; Y = 1) \\
&+ 3 \times 1 \times P\,(X = 3; Y = 1) + 4 \times 1 \times P\,(X = 4; Y = 1) + \ldots + \\
&+ 4 \times 3 \times P\,(X = 4; Y = 3) = 4.07 \\
cov\,(X, Y) &= E\,(XY) - E\,(X)\,E\,(Y) = 4.07 - 2.25 \times 1.94 = -0.295
\end{aligned}$$

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.

4.3.1 Propriedades da covariância

Proposição 4.5 *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, por isso, 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.

Para tal, propomos o coeficiente de correlação,

Definição 4.7 *Se* (X, Y) *é um par aleatório, define-se* ***coeficiente de correlação*** *de* (X, Y) *por*

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

4.3.2 Propriedades do coeficiente de correlação

Proposição 4.6 *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 muito importante: O coeficiente de correlação só permite quantificar a relação entre duas v.a.'s, desde que se admita que essa relação é linear.

Por exemplo, se via coeficente de correlação, quisermos avaliar a relação linear:

- $Y = a + bX^2$, devemos determinar $\rho\left(Y, X^2\right)$;
- $\cos Y = a + bX$, devemos determinar $\rho(\cos Y, X)$.

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

Exemplo 4.13 *Voltando aos exemplos 4.10 e 4.11,*

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

Assim é difícil de admitir uma relação linear entre X e Y.

4.4 Outros valores esperados sobre um par aleatório

Definição 4.8 *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*

- $E[\psi(X,Y)] = \displaystyle\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\psi(x_i,y_j)\,P(X = x_i; Y = y_j)$*, caso o par seja discreto com valores no suporte* $S_{(X,Y)} = \{(x_1,y_1),(x_2,y_2),\ldots\}$*;*
- $E[\psi(X,Y)] = \displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\psi(x,y)\,f_{(X,Y)}(x,y)\,dx\,dy$*, caso o par seja absolutamente contínuo com função densidade de probabilidade conjunta* $f_{(X,Y)}$.

(desde que existam)

4.5 Outros parâmetros de localização de uma variável aleatória

4.5.1 A mediana

Definição 4.9 *Seja X uma v.a. absolutamente contínua. Designa-se por **mediana** de X, o valor m_e de X que verifica,*

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

Exemplo 4.14 *Para a v.a. contínua apresentada no exemplo 2.7, 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 que garanta que $P(X \leq m_e)$ seja exactamente igual a $1/2$. Daí segue-se uma diferente definição de mediana para este tipo de v.a.'s.

Definição 4.10 *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 4.15 *Para a v.a. discreta Y, apresentada no exemplo 4.1, a mediana tem o valor $m_e = 3$ porque $P(Y \leq 2) = 0.45 < 0.5$ e $P(Y \leq 3) = 0.75 \geq 0.5$.*

4.5.2 O quantil

Função inversa

Definição 4.11 *Seja F a função distribuição de uma variável aleatória. Define-se a sua inversa por,*

$$\overleftarrow{F}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}, \quad 0 \leq u \leq 1$$

Nota: Se F é a função distribuição de uma v.a. absolutamente contínua, então $\overleftarrow{F}(u)$ é o valor x que satisfaz a igualdade,

$$F(x) = u, \quad u \in [0,1]$$

Exemplo 4.16 *A função inversa da função distribuição*

Exemplo 4.17 *A função inversa da função distribuição*

$$F(x) = \begin{cases} 0, & x \leq 100 \\ 1 - \frac{100}{x}, & x > 100 \end{cases}$$

é

$$\overleftarrow{F}(u) = \begin{cases} 100, & u = 0 \\ \frac{100}{1-u}, & 0 < u < 1 \\ +\infty, & u = 1 \end{cases}$$

A função inversa da função distribuição

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.25, & 0 \leq x < 1 \\ 0.60, & 1 \leq x < 2 \\ 0.90, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

é

$$\overleftarrow{F}(u) = \begin{cases} 0, & 0 \leq u \leq 0.25 \\ 1, & 0.25 < u \leq 0.60 \\ 2, & 0.6 < u \leq 0.90 \\ 3, & u > 0.90 \end{cases}$$

Definição de quantil

Considere-se um valor de probabilidade p, com $p \in]0,1[$.

Definição 4.12 *Seja X uma v.a. absolutamente contínua. Designa-se por **quantil** de probabilidade p da v.a. X, o valor x_p de X que verifica,*

$$P(X \leq x_p) = p, \quad p \in]0,1[$$

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

$$x_p = \overleftarrow{F}(p), \quad p \in]0,1[$$

Exemplo 4.18 *Para a v.a. contínua apresentada no exemplo 2.7, o quantil de probabilidade 0.9 de X terá o valor $x_{0.9} = 1000$, porque*

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

Definição 4.13 *Seja X uma v.a. discreta. O **quantil** de probabilidade p de X, que representamos por x_p, é o menor valor de X que verifica,*

$$P(X \leq x_p) \geq p, \quad p \in]0,1[$$

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

$$x_p = \overleftarrow{F}(p), \quad p \in]0,1[$$

Exemplo 4.19 *Para a v.a. discreta Y, apresentada no exemplo 4.1, o quantil de probabilidade 0.8 tem o valor* $y_{0.8} = 4$, *porque* $P(Y \leq 3) = 0.75 < 0.8$ *e* $P(Y \leq 4) = 0.9 \geq 0.8$.

Nota: Alguns autores representam o quantil de probabilidade p de uma v.a. X com função distribuição F, por $\overleftarrow{F}(p)$, querendo dizer que

$$P\left(X \leq \overleftarrow{F}(p)\right) = p \quad \text{ou seja} \quad F\left(\overleftarrow{F}(p)\right) = p$$

4.5.3 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 4.14 *A* ***moda*** *da v.a. X, é o valor m_o tal que*

a) $\max\limits_{x_i \in S_X} P(X = x_i) = P(X = m_o)$, *caso X seja uma v.a. discreta com valores no suporte* $S_X = \{x_1, x_2, \ldots\}$;

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

Exemplo 4.20 *Para a v.a. discreta Y do exemplo 4.1, a moda tem o valor $m_o = 2$.*
Para a v.a. contínua descrita no exemplo 2.7 a moda tem o valor $m_o = 100$.

Nota: Caso haja mais do que um valor da v.a. em que ocorra um máximo de probabilidade, dizemos que a moda não existe. (É o que acontece com a v.a. X descrita no exemplo 4.2).

4.6 Outros parâmetros de dispersão de uma variável aleatória

4.6.1 A distância média

Definição 4.15 *Seja X uma v.a.. Define-se o* ***desvio médio*** *de X por,*

$$E\left(|X - E(X)|\right)$$

caso exista.

Capítulo 5

DISTRIBUIÇÕES DISCRETAS IMPORTANTES

5.1 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:

$$X \text{ - n}^{\underline{o}} \text{ de peixes saudáveis na amostra extraída de 3 peixes}$$

Pretendemos deduzir a função de probabilidade desta v.a.

Estando apenas em causa o n.º de peixes saudáveis, a ordem por que são extraídos é irrelevante. Assim podemos considerar que o resultado da experiência são conjuntos de 3 peixes. Com esta abordagem os diferentes conjuntos que podemos 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 de entre os 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 $\{S, S, D\}$, isto é, o total de conjuntos em que devem figurar 2 peixes saudáveis seleccionáveis de entre os 5 peixes saudáveis que existem no aquário e 1 peixe doente a seleccionar de entre os $9 - 5$ peixes doentes (não saudáveis) que existem no aquário. Assim, o número de casos favoráveis à ocorrência do acontecimento $\{X = 2\}$ é,

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

Então

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

Repetindo o mesmo tipo de raciocínio para os outros valores da v.a. X, a função de probabilidade da v.a. X também pode ser expressa por

$$X\begin{cases} 0 & 1 & 2 & 3 \\ \\ \frac{\binom{5}{0}\binom{9-5}{3-0}}{\binom{9}{3}} & \frac{\binom{5}{1}\binom{9-5}{3-1}}{\binom{9}{3}} & \frac{\binom{5}{2}\binom{9-5}{3-2}}{\binom{9}{3}} & \frac{\binom{5}{3}\binom{9-5}{3-3}}{\binom{9}{3}} \end{cases}$$

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

5.1.1 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 desta 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:

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

e diz-se ter distribuição Hipergeométrica de parâmetros (N,M,n).

Obviamente $N,M,n\in\mathbb{N}$, $M\leq N$ 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 declaração "X tem distribuição Hipergeométrica de parâmetros (N,M,n)", pode ser escrita abreviadamente $X\sim H\,(N,M,n)$.

5.1.2 Coeficientes importantes

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

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**.

5.2 Distribuição Binomial

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 **com 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

Pretendemos deduzir a função de probabilidade desta v.a.

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 este resulta de ter sido extraída uma das amostras do conjunto, $\{(S,S,D), (S,D,S), (D,S,S)\}$, pelo que

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

Ora, tendo em conta o método de extracção com reposição

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

De novo o facto 3 corresponde a $\binom{3}{2}$ por ser o total de amostras com dois peixes saudáveis elegíveis de entre os 3 peixes da amostra.

Se repetirmos este processo de cálculo de probabilidades para os restantes valores da v.a. X, obtemos-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 v.a. X diz-se ter distribuição **Binomial de parâmetros $\left(3, \frac{5}{9}\right)$** ou, abreviadamente, $\mathbf{X \sim B\left(3, \frac{5}{9}\right)}$.

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 $p = \dfrac{5}{9}$.

Observações:

- Repare que, mesmo sendo finito o nº 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 nº 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.

5.2.1 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,

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

Consideremos agora, a repetição por $n \in \mathbb{N}$ vezes da experiência, nas seguintes condições:

- Mantém-se constante o valor de $p = P\left(A\right) = P\left(\text{sucesso}\right)$ em todas as experiências;
- Os resultados de sucessivas extrações são acontecimentos independentes.

Associemos aos resultados das n experiências a v.a.

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

ou

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

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

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

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, $\mathbf{X \sim B\left(n, p\right)}$.

Coeficientes importantes

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

Observações

I. Podemos seleccionar um número infinito de elementos da população, quer esta tenha dimensão finita ou infinita;

II. 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;

III. 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”).

IV. Quando uma experiência consiste em, numa única extração de um elemento de uma população, se registar a ocorrência de “sucesso” ou de “insucesso”, designamo-la por prova de Bernoulli.
Podemos associar-lhe uma v.a. indicatriz de sucesso:

$$Y = \begin{cases} 0, & \text{se ocorre insucesso} \\ 1, & \text{se ocorre sucesso} \end{cases}$$

Se $p = P\,(\text{sucesso})$ dizemos que Y tem distribuição de Bernoulli de parâmetro p, $Y \sim Ber\,(p)$, ou ainda $Y \sim B\,(1, p)$.

Exercício: Prove que $E\,(Y) = p$ e que $V\,(Y) = p\,(1-p)$

V. 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* realizarem-se independentemente de prova para prova;
- A probabilidade de “sucesso” manter-se constante em todas as n *provas de Bernoulli.*

VI. Numa sucessão de $n \in \mathbb{N}$ provas de Bernoulli com a mesma probabilidade de “sucesso”, p, e sendo $Y_1, \ldots, Y_n$ as respectivas v.a.'s indicatrizes de sucesso, o total X de ”sucessos”nessas provas,

$$X = \sum_{i=1}^{n} Y_i$$

tem distribuição Binomial de parâmetros (n, p), isto é, $X \sim B\,(n, p)$.

- Este resultado permite a demonstração simples de que, se $X \sim B\,(n, p)$, então $E\,(X) = np$ e $V\,(X) = np\,(1-p)$

  Efectivamente, pelas propriedades gerais do valor médio (ver proposição 4.1) e da variância (ver proposição 4.3)

$$\begin{aligned} E\,(X) &= E\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} E\,(Y_i) = \sum_{i=1}^{n} p = np \\ V\,(X) &= V\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} V\,(Y_i) = \sum_{i=1}^{n} p\,(1-p) = np\,(1-p) \end{aligned}$$

A seguir, apresentamos uma demonstração deste resultado.

Proposição 5.1 *Seja Y uma v.a. com distribuição $B\,(1, p)$ e X uma v.a. com distribuição $B\,(n, p)$, com $n \in \mathbb{N}$ e $0 < p < 1$.*
Se X e Y são v.a.'s independentes, então a v.a. $T = X + Y$ tem distribuição $B\,(n+1, p)$.

Demonstração: O suporte da v.a. T é $S_T = \{0, 1, \ldots, n+1\}$

Para $k \in S_T$,

$$\begin{aligned} P(T=k) &= P(Y=0; X=k) + P(Y=1; X=k-1) \\ &= (1-p)\binom{n}{k}p^k(1-p)^{n-k} + p\binom{n}{k-1}p^{k-1}(1-p)^{n+1-k} \\ &= \underbrace{\left[\binom{n}{k} + \binom{n}{k-1}\right]}_{\binom{n+1}{k}} p^k (1-p)^{n+1-k} \\ &= \binom{n+1}{k} p^k (1-p)^{n+1-k} \end{aligned}$$

Podemos agora apresentar o seguinte teorema de importância relevante:

Teorema 5.1 *Se $X_1, X_2, \ldots, X_k$ são v.a.'s* ***independentes*** *tais que $X_i \sim B(n_i, p)$, $\quad i = 1, \ldots, k$, então*

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

Nota importante: Realçamos a característica da probabilidade p de sucesso ser igual para todas as v.a.'s $X_1, X_2, \ldots, X_k$.

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

p -> a probabilidade tem que ser sempre constante

5.3 Distribuição de Poisson

Exemplo 5.1 *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 a indemnização a 12 dos 10000 segurados com apólices que cobrem este tipo de acidente.

Defina-se a v.a. X = nº de segurados que morrem num ano, de entre os 10000 segurados.

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 devidas ao elevado valor de $\binom{10000}{12}$ e também devidas 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, seja 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 a probabilidade

$$P(X_n = k), \;\; k \in \; S_{X_n} = \{0, 1, 2, \ldots, n\},$$

para valores de n a convergir para $+\infty$?

Se considerarmos $p_n = \lambda/n$,

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

Ora

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

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

e portanto

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

Este resultado permite a apresentação de uma nova distribuição, nomeadamente:

função probabilidade de uma distribuição poisson

I. Se considerarmos uma v.a. Y que regista o total de sucessos observados numa sucessão infinita de experiências de Bernoulli e a sua função de probabilidade for

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

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

Exercício 5.1 *Verificar que se trata de facto de uma função de probabilidade.*

Resolução: É importante saber ou recordar o seguinte resultado: $e^x=\sum_{i=0}^{+\infty}\frac{x^n}{n!},\quad \forall\, x\in\mathbb{R}.$

O suporte da v.a. Y *é* $S_Y=\mathbb{N}\cup\{0\}$

- $P(Y=k)=e^{-\lambda}\frac{\lambda^k}{k!}\geq 0,\ \forall\, k\in S_Y$
- $\sum_{k\in S_Y}P(Y=k)=1\Leftrightarrow\sum_{k=0}^{+\infty}e^{-\lambda}\frac{\lambda^k}{k!}=1\Leftrightarrow e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!}=1\Leftrightarrow e^{-\lambda}e^{\lambda}=1\Leftrightarrow 1=1$

Coeficientes importantes

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

Aproximação da distribuição Binomial pela distribuição de Poisson

II. 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 é, em determinadas condições, a distribuição binomial pode ser aproximada pela distribuição de Poisson.

Regra 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 evidentemente tem distribuição $B(n,(1-p))$

Exemplo 5.2 *Se aplicarmos esta aproximação ao exemplo 5.1, 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* $n = 40 \geq 30$*,* $np = 36 > 5$ *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 5.2 *Se* $Y_1, Y_2, \ldots, Y_m$ *são v.a.'s independentes tais que* $Y_i \sim P(\lambda_i)\,,\quad i = 1, \ldots, m$*, então*

$$Y_1 + Y_2 + \ldots + Y_m \sim P(\lambda_1 + \lambda_2 + \ldots + \lambda_m)$$

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 nº de automóveis que chegam em 15 minutos é Poisson com parâmetro $\lambda = 180/4 = 45$ automóveis. Por sua vez a distribuição do nº de automóveis que chegam durante o período do almoço é Poisson de parâmetro $\lambda = 2 \times 180 = 360$ automóveis.

5.3.1 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 $t \in \mathbb{R}_0^+$ unidades de observação.

$\{N_t\}_{t \in \mathbb{R}_0^+}$, diz-se um **processo de Poisson** de intensidade (ou parâmetro) $\beta \in \mathbb{R}^+$, se

- $N(0) = 0$;

- Para qualquer intervalo de $h \in \mathbb{R}^+$ unidades de observação, N_t tem distribuição de Poisson com parâmetro βh;
- Em intervalos de observação disjuntos, o n.º de sucessos neles observados são v.a.'s independentes;
- Para um intervalo de unidades de observação $]t, t+h]$, $t, h \in \mathbb{R}^+$, o n.º de sucessos, $N_{t+h} - N_t$, tem distribuição de Poisson com parâmetro βh.

A intensidade β pode ser interpretada como a taxa média de ocorrência de sucessos por **unidade de observação**.

Exemplo 5.3 *Num processo de fabricação de placas de vidro surgem 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 intensidade $\beta = 0.4$ bolhas/m^2.*

- *A probabilidade de, numa placa com 4.5 m^2, haver pelo menos uma bolha, é*

$$P(N_{4.5} \geq 1) = 1 - P(N_{4.5} = 0) = 1 - e^{-1.8} \frac{(1.8)^0}{0!} = 0.834701$$

  porque $N_{4.5}$ tem distribuição de Poisson com parâmetro $0.4 \times 4.5 = 1.8$.

- *Em média, observar-se-ão 2.4 bolhas em placas com 6 m^2, porque $N_6 \sim P(0.4 \times 6) \equiv P(2.4)$ e então*

$$E(N_6) = 2.4.$$

- *Numa placa com 4.5 m^2 e noutra placa com 6 m^2, o total de bolhas observáveis nestas duas placas tem distribuição $P(4.2)$ porque:*
  - *$N_{4.5}$ - n.º de bolhas na placa com 4.5 m^2, tem distribuição $P(1.8)$;*
  - *N_6 - n.º de bolhas na placa com 6 m^2, tem distribuição $P(2.4)$;*
  - *$N_{4.5}$ e N_6 são v.a.'s independentes;*
  - *Pelo teorema 5.2, $N_{4.5} + N_6 \sim P(1.8 + 2.4) \equiv P(4.2)$*

- *Se numa placa com 6 m^2 se registaram 8 bolhas, a probabilidade de num seu sector desta placa e com 1.5 m^2 se virem a registar 3 bolhas tem valor 0.207641602.*
  - *N_6 - n.º de bolhas na placa com 6 m^2, tem distribuição $P(2.4)$;*
  - *$N_{1.5}$ - n.º de bolhas no sector com 1.5 m^2, tem distribuição $P(1.5 \times 0.4) \equiv P(0.6)$;*
  - *$N_{1.5}$ e $N_6 - N_{1.5}$ são v.a.'s independentes, porque se referem a sectores diferentes e disjuntos;*
  - *$N_6 - N_{1.5} \sim P((6 - 1.5) \times 0.4) \equiv P(1.8)$*
  - $P(N_{1.5} = 3 \,|N_6 = 8) = \dfrac{P[(N_{1.5} = 3) \cap (N_6 = 8)]}{P(N_6 = 8)} =$

$$= \frac{P(N_{1.5} = 3; N_6 - N_{1.5} = 5)}{P(N_6 = 8)} = \frac{P(N_{1.5} = 3) \times P(N_6 - N_{1.5} = 5)}{P(N_6 = 8)} =$$

$$= \frac{e^{-0.6}\dfrac{0.6^3}{3!} \; e^{-1.8}\dfrac{1.8^5}{5!}}{e^{-2.4}\dfrac{2.4^8}{8!}} = \binom{8}{3} 0.25^3 \, 0.75^5 = 0.207641602$$

5.4 Distribuição Geométrica

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

Associemos a esta experiência a v.a.

X - **nº** de peixes extraídos até se observar **pela primeira vez** um peixe saudável.

Na dedução da função de probabilidade da v.a. X, começamos por constatar que os valores admissíveis são os valores do suporte $S_X = \mathbb{N}$.

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

Apenas para ilustração, apresentamos as seguintes determinações:

$$\begin{aligned}
P(X=1) &= P(S) = p \\
P(X=2) &= P\left(\overline{S} \cap S\right) = P\left(\overline{S}\right) P(S) = p\,(1-p) \\
P(X=3) &= P\left(\overline{S} \cap \overline{S} \cap S\right) = \left[P\left(\overline{S}\right)\right]^{3-1} P(S) = p\,(1-p)^{3-1}
\end{aligned}$$

Generalizando (e podendo ser demonstrado pelo método de indução em $k \in \mathbb{N}$).

$$P(X=k) = p\,(1-p)^{k-1}, \quad \forall\, k \in \mathbb{N}$$

Características gerais da distribuição Geométrica

Ao realizarmos uma experiência aleatória, estamos apenas interessados em verificar se um determinado acontecimento A se realiza (*sucesso*) ou não se realiza (*insucesso*) (Prova de Bernoulli).

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,

$$p = P(A) = P(\text{sucesso}) \qquad \text{com } 0 < p < 1$$

Consideremos agora, repetições **independentes** da experiência, implicando isto que se mantém constante o valor de p.

A v.a.

X - **nº** de experiências a realizar até se observar um sucesso **pela primeira vez**

tem uma função de probabilidade caracterizada por:

- Suporte $S_X = \mathbb{N}$;
- $P(X=k) = p\,(1-p)^{k-1}, \quad k \in S_X$

e diz-se ter distribuição **Geométrica de parâmetro** p.

A frase "X tem distribuição geométrica de parâmetro p", pode ser escrita de modo abreviado, $\mathbf{X \sim G\,(p)}$.

Exercício 5.2 *Verificar que é de facto uma função de probabilidade.*

Resolução: É importante saber ou recordar o seguinte resultado:

$$\textit{Se } r \textit{ é um real positivo e } m \in \mathbb{N}\ , \quad \sum_{i=0}^{m} r^i = \frac{1-r^{m+1}}{1-r}$$

O suporte da v.a. X é $S_X = \mathbb{N}$

- $P\,(X=k) = p\,(1-p)^{k-1} \geq 0, \ \forall\, k \in S_X$ *porque* $0 < p < 1$;
- $\displaystyle\sum_{k \in S_X} P\,(X=k) = 1 \Leftrightarrow \sum_{k=1}^{+\infty} p\,(1-p)^{k-1} = 1 \Leftrightarrow p \sum_{k=1}^{+\infty} (1-p)^{k-1} = 1 \Leftrightarrow p\frac{1}{1-(1-p)} = 1 \Leftrightarrow 1 = 1$

Observações

- 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 geométrica, o n.º de sucessos é fixo e igual a 1 e, o que contamos são o n.º de provas de Bernoulli a realizar de modo a se conseguir observar esse sucesso.

Coeficientes importantes

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

Função distribuição

Proposição 5.2 *Se X é uma v.a. com distribuição Geométrica de parâmetro p, a sua função distribuição é:*

$$F_X\,(x) = \begin{cases} 0, & x \in \left]-\infty, 1\right[\\ 1-(1-p)^{[x]}, & x \in \left[1, +\infty\right[\end{cases}$$

representando $[x]$ a parte inteira de x.

Demonstração: Para a demonstração, precisamos de recordar o seguinte resultado:

$$\text{Se } r \text{ é um real positivo e } m \in \mathbb{N}\,, \quad \sum_{i=0}^{m} r^i = \frac{1-r^{m+1}}{1-r}$$

- Para $x < 1$, $\quad F_X(x) = P(X \le x) = P(\emptyset) = 0$;
- Para $x \ge 1$, $\quad F_X(x) = P(X \le x) = P(X \le [x]) = \sum_{k=1}^{[x]} p(1-p)^{k-1} = p \sum_{i=0}^{[x]-1} (1-p)^i =$

$$= p\,\frac{1-(1-p)^{[x]-1+1}}{1-(1-p)} = 1-(1-p)^{[x]}$$

Propriedade da falta de memória

Proposição 5.3 *Se X é uma v.a. com distribuição $G(p)$, então*

$$P(X > s+t \,|X > s\,) = P(X > t)\,, \;\forall\, s,t \in \mathbb{N}$$

Demonstração: $\forall\, s,t \in \mathbb{N}$

$$\begin{aligned} P(X > s+t\,|X > s\,) &= \frac{P[(X > s+t) \cap (X > s)]}{P(X > s)} = \frac{P(X > s+t)}{P(X > s)} = \frac{1-F_X(s+t)}{1-F_X(s)} = \\ &= \frac{(1-p)^{s+t}}{(1-p)^s} = (1-p)^t = P(X > t) \end{aligned}$$

Exemplo 5.4 *Uma máquina produz peças e, independentemente da peça, a probabilidade de sair defeituosa tem valor $p = 0.02$. Considere a v.a. X-n.º de peças produzidas até se observar uma defeituosa.*

- *A v.a. X tem distribuição Geométrica com parâmetro* 0.02, $\quad X \sim G(0.02)$.
- *O n.º esperado de peças a produzir até se registar uma peça defeituosa é*

  $$E(X) = \frac{1}{0.02} = 50 \text{ peças}$$

  e o desvio padrão de X é

  $$\sigma(X) = \sqrt{V(X)} = \sqrt{\frac{1-0.02}{0.02^2}} \approx 49.5 \text{ peças.}$$

  Interpretação dos valores $E(X)$ e $\sigma(X)$ recorrendo à desigualdade de Chebychev

  Com probabilidade superior ou igual a 0.75, o n.º de peças a produzir até se observar uma defeituosa, tem valores no intervalo

$$\left]E\left(X\right)-2\sigma\left(X\right),E\left(X\right)-2\sigma\left(X\right)\right[\approx\left]-48.99,148.99\right[\text{ peças}$$

Tendo em conta o suporte de X*, o intervalo é:* $]1,148[$ *peças a produzir até ocorrer uma peça defeituosa.*

- *A probabilidade de se vir a registar, pela primeira vez, uma peça defeituosa após a produção de 25 peças é*

  $$P\left(X=25\right)=0.02\left(1-0.02\right)^{24}=0.012315607$$

- *A probabilidade* $P\left(46\leq X\leq 102\right)$ *tem valor*

  $$P\left(46\leq X\leq 102\right)=\sum_{k=46}^{102}P\left(X=k\right)=\sum_{k=46}^{102}0.02\left(1-0.02\right)^{k-1}$$

  sendo mais vantajoso recorrer à função distribuição para o respectivo cálculo

  $$P\left(46\leq X\leq 102\right)=P\left(X\leq 102\right)-P\left(X<46\right)=P\left(X\leq 102\right)-P\left(X\leq 45\right)=$$

  $$=F_X\left(102\right)-F_X\left(45\right)=1-\left(1-0.02\right)^{102}-\left[1-\left(1-0.02\right)^{45}\right]=$$
  $$=0.98^{45}-0.98^{102}=0.275510043$$

- *Sabendo que após a produção de 100 peças, não se registou uma defeituosa, a probabilidade de ser necessário produzir mais de 104 peças para se registar uma defeituosa é:*

  $$P\left(X>104\,|X>100\right)=P\left(X>4\right)=1-P\left(X\leq 4\right)=1-F_X\left(4\right)=\left(1-0.02\right)^4=0.92236816$$

Capítulo 6

DISTRIBUIÇÕES CONTÍNUAS IMPORTANTES

6.1 Distribuição Uniforme

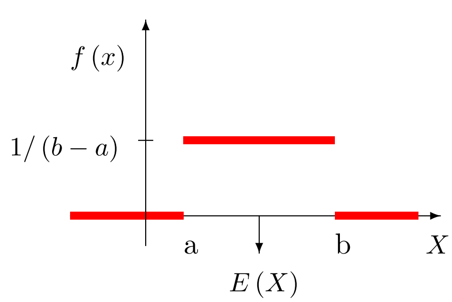
Esta distribuição aplica-se quando os valores de uma v.a. X, absolutamente contínua, podem ocorrer dentro de um intervalo de valores reais, limitado (aberto, fechado ou semi-aberto) de extremos a, b, com $a < b$, e quaisquer dois sub-intervalos com a mesma amplitude têm a mesma probabilidade de ocorrência.

Diz-se então que X tem distribuição **Uniforme** com parâmetros (a, b) (abreviadamente, escreve-se $\mathbf{X \sim U(a, b)}$).

A sua função densidade de probabilidade é

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

Figura 6.1: Função densidade da distribuição Uniforme



6.1.1 Coeficientes importantes

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

Exercício 6.1 *Mostre que, se X é uma v.a. absolutamente contínua e com distribuição $U(a,b)$, então:*

$$E(X) = \frac{a+b}{2}, \quad E\left(X^2\right) = \frac{b^2+ab+a^2}{3} \quad e \quad V(X) = \frac{(b-a)^2}{12}$$

Exemplo 6.1 *No exemplo 4.2 utilizámos esta distribuição para descrever o peso perdido com o primeiro tipo de dieta. X tem distribuição uniforme no intervalo $[2,4]$.*

Também $E(X) = \frac{2+4}{2} = 3$ *kg e* $\sigma(X) = \sqrt{V(X)} = \sqrt{\frac{(4-2)^2}{12}} = \sqrt{\frac{1}{3}}$

6.1.2 Função distribuição

$$F_X(x) \equiv P(X \leq x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$$

Teorema 6.1 *Seja X uma v.a. com função distribuição F. A v.a. $U = F(X)$ tem distribuição Uniforme com parâmetros $(0,1)$, isto é, $F(X) \sim U(0,1)$.*

Demonstração:
Para $u \in [0,1]$ e atendendo à definição da inversa de uma função distribuição (ver a definição 4.11)

$$F_U(u) = P(U \leq u) = P\left(\overleftarrow{F}(X) \leq u\right) = P\left(X \leq \overleftarrow{F}(X)(u)\right) = F\left(\overleftarrow{F}(u)\right) = u$$

6.2 Distribuição Exponencial

A distribuição 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 a v.a. X verifica a propriedade da falta de memória:

$$P\left(X > s + t\,|X > s\,\right) = P\left(X > t\right),\; \forall\, s, t \in \mathbb{R}^{+} \tag{6.2.1}$$

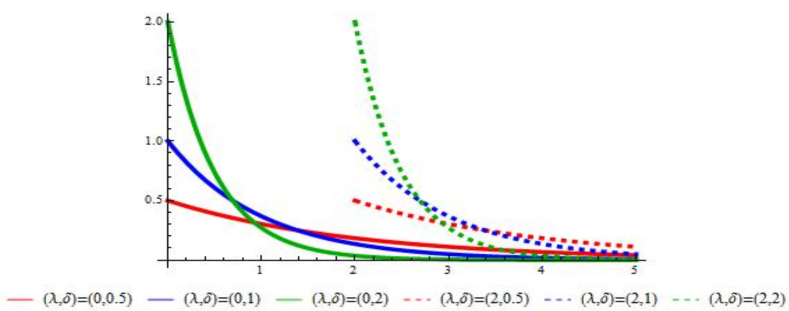
Definição 6.1 *Uma v.a. X, absolutamente contínua, diz-se ter **Exponencial** com parâmetros (λ, δ), se a sua função densidade de probabilidade for*

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

e escreve-se de modo abreviado, $\mathbf{X} \sim \mathbf{E}\left(\lambda, \delta\right)$.

Os parâmetros λ e δ deverão ser valores reais e $\delta > 0$.

Figura 6.2: Exemplos de funções densidade da distribuição exponencial



Propriedade da falta memória: A propriedade de não memória expressa na equação (6.2.1), só é válida quando $\lambda = 0$.

Mas agora, expressamos matematicamente e de modo mais abrangente esta propriedade.

- Se $X \sim E\left(0, \delta\right)$, então $P\left(X > x + h\,|X > x\,\right) = P\left(X > h\right), \quad \forall\, x, h \in \mathbb{R}^{+}$
- Se $X \sim E\left(\lambda, \delta\right)$, então $P\left(X > x + h\,|X > x\,\right) = P\left(X > h + \lambda\right), \quad \forall\, x \in \left]\lambda, +\infty\right], \forall\, h \in \mathbb{R}^{+}$

6.2.1 Coeficientes importantes

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

6.2.2 Função distribuição

$$F_X(x) = \begin{cases} 0, & x < \lambda \\ 1 - e^{-\frac{x-\lambda}{\delta}}, & x \geq \lambda \end{cases}$$

Exemplo 6.2 *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 =nº 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,

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

Teorema 6.2 *Se Y é uma v.a. com distribuição $E(0,1)$, então a v.a. $X = \lambda + \delta Y$ tem distribuição $E(\lambda, \delta)$.*

6.2.3 Relação entre a distribuição Exponencial e o Processo de Poisson

Admitamos que $\{N_t\}_{t \geq 0}$ é um **Processo de Poisson de intensidade β** em que, para um 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. absolutamente contínua e os seus valores admissíveis são os reais superiores ou iguais a 0 (o suporte da T é $S_T = \mathbb{R}_0^+$.)

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

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

Tendo em conta a nota 4 da secção 2.3 (Variável aleatória absolutamente contínua), é possível deduzir a função densidade de probabilidade da v.a. T.

Qualquer que seja $t \in \mathbb{R}^+$,

$$\begin{aligned}
\lim_{h\to 0}\frac{P(t<T\le t+h)}{h} &= \lim_{h\to 0}\frac{P(T>t)-P(T>t+h)}{h}\\
&= \lim_{h\to 0}\frac{e^{-\beta t}-e^{-\beta(t+h)}}{h} = e^{-\beta t}\lim_{h\to 0}\frac{1-e^{\beta h}}{h}\\
&= e^{-\beta t}\lim_{h\to 0}\beta e^{\beta h} = \beta e^{-\beta t} \equiv f_T(t)
\end{aligned}$$

Assim sendo, a função densidade de probabilidade da v.a. T é:

$$f_T(t) = \begin{cases} 0 & t<0 \\ \beta e^{-\beta t} & t \ge 0 \end{cases}$$

que reconhecemos ser a função densidade de probabilidade de uma distribuição **Exponencial** com **parâmetros** $\left(\mathbf{0}, \dfrac{\mathbf{1}}{\beta}\right)$.

Resumindo:

Se $\{Nt\}_{t\ge 0}$ é um **Processo de Poisson de intensidade** 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** $(\mathbf{0}, \mathbf{1}/\beta)$.

6.3 Distribuição de Weibull

A distribuição de Weibull, devida a Waloddi Weibull (1951), é uma altermativa à distribuição Exponencial quando não se garante a propriedade de falta de memória (ver equação 6.2.1) que caracteriza a distribuição Exponencial. O campo de aplicações da distribuição de Weibull é vasto, abrangendo áreas científicas como, por exemplo, a física, a biologia, o ambiente, a fiabilidade e a engenharia de confiabilidade.

Definição 6.2 *Uma v.a. X, absolutamente contínua, tem distribuição de Weibull com parâmetros $(\lambda, \delta, \gamma)$, se a sua função densidade é:*

$$f_X(x) = \begin{cases} 0, & x<\lambda \\ \dfrac{\gamma}{\delta}\left(\dfrac{x-\lambda}{\delta}\right)^{\gamma-1} e^{-\left(\dfrac{x-\lambda}{\delta}\right)^{\gamma}}, & x \ge \lambda \end{cases} \qquad \lambda\in\mathbb{R},\ \delta\in\mathbb{R}^+,\ \gamma\in\mathbb{R}_0^+$$

Abreviadamente escrevemos $X \sim W(\lambda,\delta,\gamma)$.

6.3.1 Função distribuição

$$F_X(x) = P(X\le x) = \begin{cases} 0, & x<\lambda \\ 1-e^{-\left(\frac{x-\lambda}{\delta}\right)^{\gamma}}, & x\ge\lambda \end{cases}$$

6.3.2 Coeficientes importantes

$$E(X) = \lambda + \delta\,\Gamma\left(1 + \frac{1}{\gamma}\right) \qquad V(X) = \delta^2\left[\Gamma\left(1 + \frac{2}{\gamma}\right) - \left(\Gamma\left(1 + \frac{1}{\gamma}\right)\right)^2\right]$$

6.3.3 Notas

1. Se $\gamma = 1$, X tem distribuição Exponencial com parâmetros (λ, δ).
2. Interpretação do papel do parâmetro γ.

   Considere-se $\lambda = 0$, isto é, $X \sim W(0, \delta, \gamma)$

   Se X representar o tempo decorrido até ocorrer uma falha num equipamento, seguem-se as interpretações relativas ao parâmetro γ:

   a) Se $\gamma < 1$, alta taxa de falha no início. Esse é um comportamento típico de processos industriais em que a maioria das falhas ocorre no ínicio do processo de produção dos items ou quando a taxa de falha diminui ao se eliminar a população defeituosa de dispositivos;
   b) Se $\gamma = 1$, chance de falha independente do tempo e comportamento exponencialmente decrescente da distribuição. Processos "sem memória"em que as falhas ocorrem devido a razões aleatórias (distribuição Exponencial);
   c) Se $\gamma > 1$, chance de falha crescente com o tempo. Casos em que há um processo de envelhecimento.

6.4 Distribuição de Pareto

Desenvolvida pelo engenheiro civil, sociólogo e economista Vilfredo Pareto, esta distribuição foi proposta para modelar observações relativos a fenómenos económicos, sociais, geofísicos, actuarias, e muitos outros. Acenta no princípio de Pareto (ou regra 80-20), segundo o qual, para um certo fenómeno aleatório, 80% dos efeitos advêm de 20% das suas causas. Por exemplo,

- 20% dos clientes de uma empresa são responsáveis por 80% das vendas;
- 80% da poluição fabril tem origem em 20% das fábricas.

Extensível a outras regras, por exemplo 91-09.

Estas regras (sem relação teórica expressível), só procuram passar a ideia de, numa relação causa versus efeito, existir em simultâneo, uma significativa assimetria e concentração.

Do ponto de vista teórico, para explicação desta assimetria e concentração, podemos propor o modelo Pareto, segundo o qual, uma v.a. X que satisfaz as condições:

- $P(X > x + h\,x \,|X > x)$ não depender de x e
- $P(X > x + h\,x \,|X > x) = (1 + h)^{-\alpha}, \quad h \in]0, +\infty[\,, \quad \alpha \in \mathbb{R}^+$

diz-se ter distribuição de Pareto

Exemplo 6.3 *A probabilidade de alguém que tem mais de 1 milhão, vir a ter mais de* $(h+1)$ *milhões, é exactamente igual à probabilidade de alguém que tem mais de 100 milhões, vir a ter mais de* $100\,(h+1)$*) milhões. Significa que o valor destas probabilidades depende apenas de* $h \in \mathbb{R}^+$*. Se a v.a.* X *representar a riqueza de um indivíduo (em milhão de euros), e* X *tiver distribuição de Pareto, então*

$$P\,(X > h+1\,|X > 1\,) = P\,(X > 100\,(h+1)\,|X > 100\,) = (1+h)^{-\alpha}$$

Façamos uma avaliação númerica, considerando $h = 0.5$ *e* $\alpha = 4$

$$P\,(X > 1.5\,|X > 1\,) = P\,(X > 150\,|X > 100\,) = (1.5)^{-4} \approx 0.198$$

Conclusão: Qualquer que seja a riqueza inicial (1 ou 100 milhões de euros), a probabilidade de vir a ter mais metade da sua riqueza incial (0.5 milhão de euros ou 50 milhões de euros), é exactamente igual.

Mas também, para uma riqueza inicial superior a, 1 milhão de euros ou 100 milhões de euros, as probabilidades de vir a ter mais 0.5 milhão de euros, são, respectivamente:

$$P\,(X > 1.5\,|X > 1\,) = (1.5)^{-4} \approx 0.198 \quad e \quad P\,(X > 100.5\,|X > 100\,) = (1.005)^{-4} \approx 0.980$$

Definição 6.3 *Uma v.a.* X *absolutamente contínua, diz-se ter distribuição Pareto com parâmetros* (δ, α)*,* $(\delta, \alpha \in \mathbb{R}^+)$ *se tiver f.d.p. e f.d., respectivamente:*

$$f_X\,(x) = \begin{cases} 0, & x < \delta \\ \dfrac{\alpha}{\delta}\left(\dfrac{x}{\delta}\right)^{-\alpha-1}, & x \geq \delta \end{cases}$$

$$F_X\,(x) = \begin{cases} 0, & x < \delta \\ 1 - \left(\dfrac{x}{\delta}\right)^{-\alpha}, & x \geq \delta \end{cases}$$

Abreviadamente escrevemos $X \sim Par\,(\delta, \alpha)$*.*

δ é um parâmetro de escala. α é um parâmetro de forma conhecido por índice de Pareto ou índice de cauda.

6.4.1 Coeficientes importantes

- $E\left(X^{\xi}\right) = \dfrac{\alpha\delta^{\xi}}{\alpha - \xi},\ \alpha > \xi$
- $E\,(X) = \dfrac{\alpha\delta}{\alpha - 1},\ \alpha > 1$
- $V\,(X) = \dfrac{\alpha\delta^2}{(\alpha-2)\,(\alpha-1)^2},\ \alpha > 2$
- Quantil de probabilidade p, $\qquad x_p = \delta\ (1-p)^{-1/\alpha}$

6.5 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 ricas e importantes que, muito métodos estatísticos só podem ser deduzidos e 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.

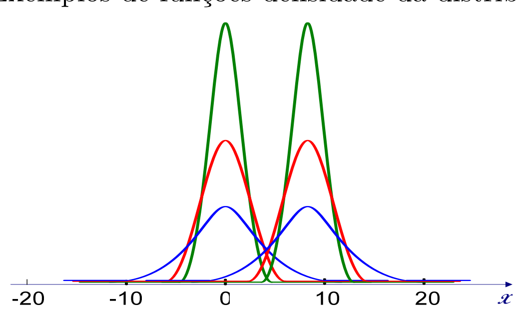
Definição 6.4 *Diz-se que a v.a. X tem distribuição* ***Normal com parâmetros*** $\left(\mu, \sigma^{\mathbf{2}}\right)$ *e escreve-se de modo abreviado,* $\mathbf{X} \sim \mathbf{N}\left(\mu, \sigma^{\mathbf{2}}\right)$*, se a sua função densidade de probabilidade for,*

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

Os parâmetros μ e σ deverão satisfazer $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}^+$.

O parâmetro μ indica o ponto de simetria da densidade e o parâmetro σ expressa a dispersão da densidade. Na figura , 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 $\mu = 4$. As curvas da menos "achatada" à mais "achatada", correspondem a $\sigma = 0.5$, $\sigma = 1$ e a $\sigma = 2$, respectivamente.

Figura 6.3: Exemplos de funções densidade da distribuição Normal



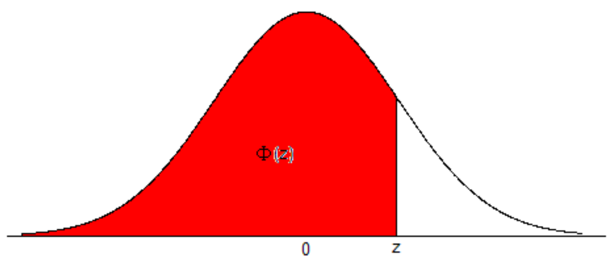
6.5.1 Distribuição Normal Reduzida

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, $\mathbf{Z} \sim \mathbf{N}\left(\mathbf{0}, \mathbf{1}\right)$. Como veremos adiante, esta distribuição tem um papel muito importante no cálculo de probabilidades de qualquer distribuição normal.

Definição 6.5 *Seja Z uma v.a. com distribuição Normal Reduzida, $Z \sim N\left(0, 1\right)$. A função de distribuição da v.a. Z define-se e denota-se por*

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

Figura 6.4: Função distribuição de Z



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

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 a valores de função de distribuição Φ, seja por uso de uma ferramenta de cálculo seja por leitura de valores editados numa tabela da função de distribuição normal reduzida.

Exemplo 6.4

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

e

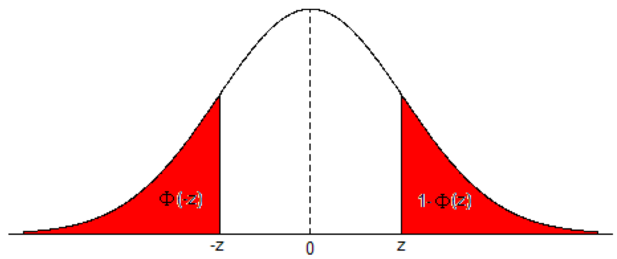
$$P\left(Z \leq -1.27\right) = ?$$

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

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

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

Figura 6.5: Consequência da simetria da densidade de Z para a sua função de distribuição



Teorema 6.3 *Se Z é uma v.a. com distribuição Normal Reduzida, $Z \sim N(0,1)$, então*

$$P(Z \leq -z) = P(Z > z), \quad \forall\, z \in \mathbb{R}$$

ou ainda,

$$\Phi(-z) = 1 - \Phi(z), \quad \forall\, z \in \mathbb{R}$$

6.5.2 Como calcular quantis de uma v.a. Normal Reduzida

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 6.5
$z_{0.2} \approx 0.84$ *porque* $P(Z \leq 0.84) = 0.7995 \approx 0.8$
$z_{0.8} \approx -0.84$ *porque* $P(Z \leq -0.84) = 1 - P(Z \leq 0.84) = 1 - 0.7995 \approx 0.2$

Coeficientes importantes

Proposição 6.1 *Se $Z \sim N(0,1)$, então*

$$E(Z) = 0 \qquad e \qquad V(Z) = 1$$

Demonstração:

- $$E(Z) = \int_{-\infty}^{+\infty} z\, f_Z(z)\, dz = \int_{-\infty}^{+\infty} z\, \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz = \left[-\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \right]_{-\infty}^{+\infty} = 0$$

- $$E\left(Z^2\right) = \int_{-\infty}^{+\infty} z^2\, f_Z(z)\, dz = \int_{-\infty}^{+\infty} z^2\, \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz =$$
$$= \left[-\frac{z}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \right]_{-\infty}^{+\infty} + \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz}_{=1} = 0 + 1$$

- $V(Z) = E(Z^2) - E^2(Z) = 1$

6.5.3 Resultados gerais para v.a.'s com distribuição Normal

Seja X uma v.a. com distribuição Normal de parâmetros μ e σ^2, $X \sim N(\mu, \sigma^2)$.
Por uma simples mudança de variável na sua função densidade da v.a. X, concluí-se que:

Teorema 6.4 *Se $X \sim N(\mu, \sigma^2)$, então*

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

Corolário 6.1 *Se $X \sim N(\mu, \sigma^2)$, então*

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

Corolário 6.2 *Se $X \sim N(\mu, \sigma^2)$ e $Z \sim N(0, 1)$, então*

$$P(X \leq a) = P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) = P\left(Z \leq \frac{a - \mu}{\sigma}\right), \quad \forall\, a \in \mathbb{R}$$

Exemplo 6.6 *Admitamos que $X \sim N(1, 4)$.*

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

O próximo teorema generaliza o anterior.

Teorema 6.5 *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

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

Finalmente um teorema de largas aplicações estatísticas só válido para a distribuição Normal.

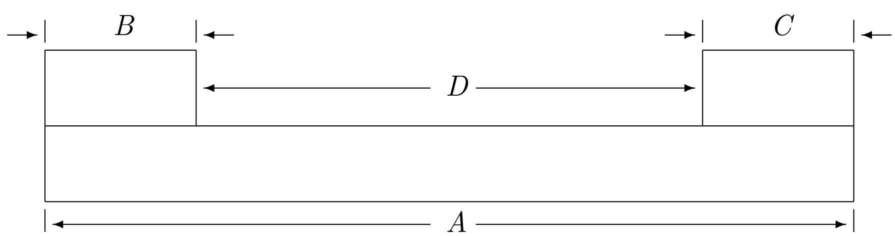
Teorema 6.6 *Sejam $X_1, X_2, \ldots, X_k$ v.a.'s independentes com distribuição $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, k$. Considerem-se $a_1, a_2, \ldots, a_k, b$ constantes reais, com algum $a_i \neq 0$. A v.a.*

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

Repare que

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

Exemplo 6.7 *Um molde de planificação*



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

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

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 $N\left(E\left(D\right), V\left(D\right)\right)$. Como

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

então $D \sim N\left(6, 0.2\right)$.

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

6.6 Distribuição do Qui-quadrado

Definição 6.6 *Uma v.a. X diz-se ter distribuição do **Qui-Quadrado** com $n \in \mathbb{N}$ graus de liberdade, (abreviadamente, $X \sim \chi_n^2$), se é v.a. absolutamente contínua com função densidade de probabilidade:*

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

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

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

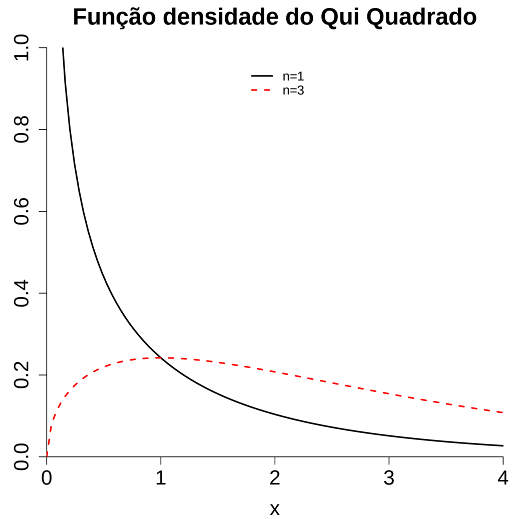


Figura 6.6: Função densidade de $X \sim \chi^2_n$.

6.6.1 Coeficientes importantes

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

Notas

1. O cálculo de probabilidades ou de quantis de probabilidade associados a variáveis aleatórias com distribuição do Qui-quadrado é feito através de tabelas desta distribuição, ou através de ferramentas de cálculo como calculadoras e computadores;

2. Denotamos por $\chi^2_{n:\alpha}$ o quantil de probabilidade $1-\alpha$ de uma v.a. X com distribuição χ^2_n. Isto é, $\chi^2_{n:\alpha}$, é o valor da v.a. X que satisfaz

$$P(X \leq \chi^2_{n:\alpha}) = 1-\alpha \qquad \text{ou} \qquad P(X > \chi^2_{n:\alpha}) = \alpha$$

Teorema 6.7 *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*

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

*tem distribuição do **Qui-quadrado** com n graus de liberdade.*

6.7 Distribuição t-Student

Definição 6.7 *Uma v.a. T diz-se ter distribuição **t-Student** com $n \in \mathbb{N}$ graus de liberdade, (abreviadamente, $T \sim t_n$), se é v.a. absolutamente contínua com função densidade de probabilidade:*

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

Apresenta-se na figura 6.7 o gráfico da função densidade de uma v.a. $T \sim t_n$.

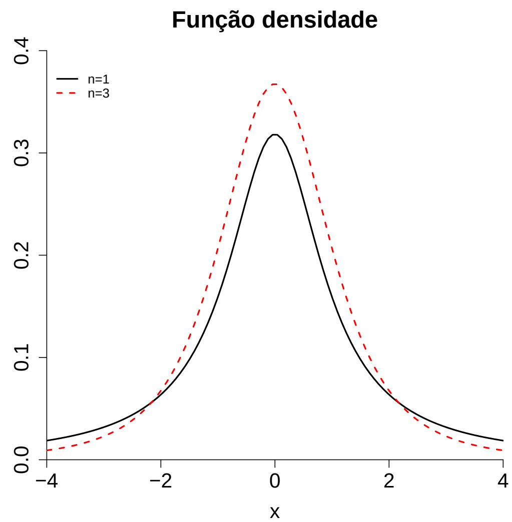


Figura 6.7: Função densidade de $T \sim t_n$.

Coeficientes importantes

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

Teorema 6.8 *Sejam $Z \sim N(0,1)$ e $Y \sim \chi^2_n$, v.a.'s independentes. A variável aleatória*

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

*tem distribuição **t-Student** com n graus de liberdade.*

6.7.1 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 pode ser feito através de tabelas da função de distribuição ou através de outras ferramentas de cálculo (calculadora, computador);

2. Denotamos por $t_{n:\alpha}$ o quantil de probabilidade $1-\alpha$ de uma v.a. T com distribuição t-Student com n graus de liberdade. Isto é, $t_{n:\alpha}$ é o valor da v.a. T que satisfaz

$$P(T \leq t_{n:\alpha}) = 1 - \alpha \qquad \text{ou} \qquad P(T > t_{n:\alpha}) = \alpha$$

3. A função de densidade da distribuição t-Student é simétrica em torno de zero. Por isso, é válida a seguinte igualdade entre quantis

$$t_{n:\alpha} = -t_{n:1-\alpha}$$

6.8 Distribuição F de Fisher

Definição 6.8 *Uma v.a. X absolutamente contínua, com função densidade de probabilidade:*

$$f_X(x) = \begin{cases} 0, & x < 0 \\ \frac{\Gamma\left(\frac{n+m}{2}\right) m^{m/2} n^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{x^{m/2-1}}{(n+m\,x)^{(m+n)/2}}, & x \geq 0 \end{cases}, \qquad m, n \in \mathbb{N}$$

*diz-se ter distribuição **F** de Fisher (ou Snedcor) com um par (m, n) de graus de liberdade.*

Abreviadamente escrevemos $X \sim F_{(m,n)}$.

Na figura 6.8 apresentamos exemplos do gráfico da função densidade de um variável aleatória com distribuição $F_{m,n}$.

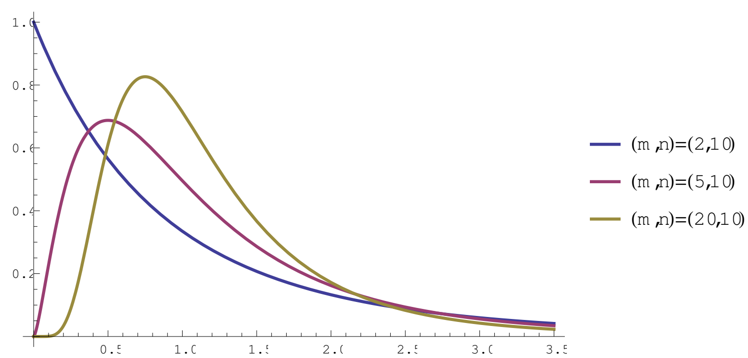


Figura 6.8: Função densidade de $X \sim F_{m,n}$.

6.8.1 Coeficientes importantes

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

Teorema 6.9 *Sejam $Y \sim \chi^2_m$ e $W \sim \chi^2_n$ v.a.'s independentes. A variável aleatória*

$$X = \frac{Y/m}{W/n} = \frac{n}{m}\frac{Y}{W}$$

*tem uma distribuição **F** (de Fisher ou Snedecor) com um par (m, n) de graus de liberdade.*

6.8.2 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 ou de outras ferramentas de cálculo como calculadoras e computadores;

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, $F_{(m,n):\alpha}$ é o valor que satisfaz

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

3. É válida a seguinte relação entre quantis

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

Capítulo 7

TEOREMA LIMITE CENTRAL (T.L.C.)

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

7.1 T.L.C. (Versão soma e Versão média de v.a.'s)

Teorema 7.1 (Teorema Limite Central (Versão soma de v.a.'s))

Seja $X_1, X_2, \ldots, X_i, \ldots$ *uma sucessão de v.a.'s independentes e identicamente distribuídas (i.i.d.).*

Admita-se que se conhece o valor médio, μ*, comum a todas as v.a.'s, assim como a variância,* σ^2*, 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,

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

ou seja, quando $n \to +\infty$*, a função distribuição da v.a.* $\frac{S_n - n\mu}{\sqrt{n}\sigma}$ *converge para a função distribuição de uma v.a. Normal Reduzida.*

Numa escrita abreviada, quando $n \to +\infty$,

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

querendo o símbolo $\overset{a}{\sim}$ *dizer "tem distribuição aproximada".*

Nota 1: Na prática, estabelece-se que, para $n \geq 30$ podemos usar o resultado do T.L.C..
Nota 2: Pelas condições exigidas às v.a.'s da sucessão $X_1, X_2, \ldots, X_i, \ldots$, concluímos que:

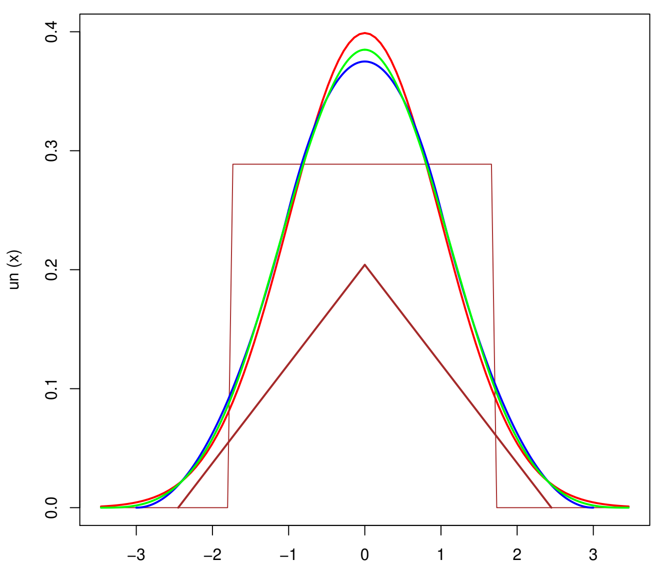
- $E(S_n) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \mu = n\mu$
- $V(S_n) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} \sigma^2 = n\sigma^2$
- Resumindo: Nas condições do T.L.C. e para $n \geq 30$, o resultado

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

pode ser expresso por:

$$\frac{S_n - E(S_n)}{\sqrt{V(S_n)}} \overset{a}{\sim} N(0,1)$$

Figura 7.1: 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 $N(0,1)$.

Exemplo 7.1 *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!"

O Teorema Limite Central tem inúmeras aplicações em Estatística, onde o estudo de médias aritméticas de v.a.'s é especialmente importante. Por isso enunciamos o T.L.C. para uma sucessão de médias aritméticas de variáveis aleatórias.

Teorema 7.2 (Teorema Limite Central (Versão média de v.a.'s))

Seja $X_1, X_2, \ldots, X_i, \ldots$ *uma sucessão de v.a.'s independentes e identicamente distribuídas (i.i.d.).*

Admita-se que se conhece o valor médio, μ, *comum a todas as v.a.'s, assim como a variância,* σ^2, *comum a todas as v.a.'s.*

Considere-se a sucessão das somas parciais $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, $n = 1, 2, \ldots$.

Então,

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

ou seja, quando $n \to +\infty$, a função distribuição da v.a. $\sqrt{n}\dfrac{\overline{X}_n - \mu}{\sigma}$ *converge para a função distribuição de uma v.a. Normal Reduzida.*

Numa escrita abreviada, quando $n \to +\infty$,

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

querendo o símbolo $\overset{a}{\sim}$ dizer "tem distribuição aproximada".

Nota 3: Estas duas versões do T.L.C. são equivalentes porque $S_n = n\overline{X}_n$, e então,

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

Nota 4: Pelas condições exigidas às v.a.'s da sucessão $X_1, X_2, \ldots, X_i, \ldots$, concluímos que:

- $E\left(\overline{X}_n\right) = E\left(\dfrac{S_n}{n}\right) = \dfrac{1}{n}E(S_n) = \dfrac{n\mu}{n} = \mu$
- $V\left(\overline{X}_n\right) = V\left(\dfrac{S_n}{n}\right) = \dfrac{1}{n^2}V(S_n) = \dfrac{n\sigma^2}{n^2} = \dfrac{\sigma^2}{n}$
- Resumindo: Nas condições do T.L.C. e para $n \geq 30$,

$$\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \overset{a}{\sim} N(0,1) \Leftrightarrow \frac{\overline{X}_n - E\left(\overline{X}_n\right)}{\sqrt{V\left(\overline{X}_n\right)}} \overset{a}{\sim} N(0,1)$$

7.2 Aplicações particulares do T.L.C.

7.2.1 Distribuição Binomial

Teorema 7.3 *Seja X uma v.a. com distribuição binomial de parâmetros (n, p). Se $n \geq 30$ e p tem valor tal que $np > 5$ e $n(1-p) > 5$, então*

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

Observação: Para justificação deste resultado, tenha em conta a observação VI da sub-secção 5.2.1.

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

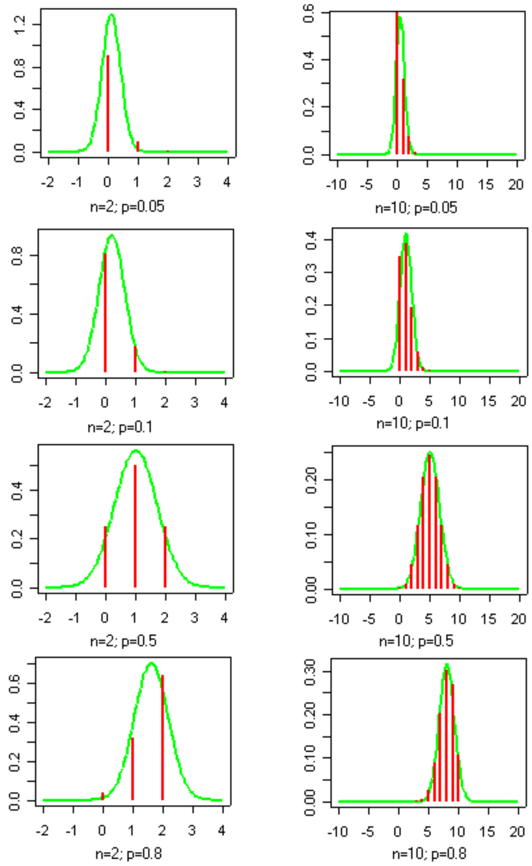
$$P(X \leq k) = P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{k - np}{\sqrt{np(1-p)}}\right) \approx P\left(Z \leq \frac{k - np}{\sqrt{np(1-p)}}\right), \ k = 0, 1, \ldots$$

Exemplo 7.2 *Suponhamos que* $X \sim B\left(100, \frac{1}{4}\right)$ *e que queremos calcular a* $P(16 \le X \le 30)$ *e a* $P(X = 27)$*.*
Como $np = \frac{100}{4} = 25 > 5$ *e* $n(1-p) = 100\frac{3}{4} = 75 > 5$,

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

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

Figura 7.2: Ilustração da aproximação da dist. Binomial pela dist. Normal para $n = 2$ e $n = 10$



7.2.2 Distribuição de Poisson

Teorema 7.4 *Seja X uma v.a. com distribuição de Poisson de parâmetro λ. Se $\lambda > 5$,*

$$\frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - \lambda}{\sqrt{\lambda}} \overset{a}{\sim} N(0,1)$$

Observação: Como aproximamos a função distribuição de uma v.a. discreta X pela distribuição distribuição de uma v.a. contínua, a aproximação deve ser feita sobre as respectivas funções distribuição, 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 7.3: Ilustração da aproximação da dist. Binomial pela dist. Normal para $n = 30$

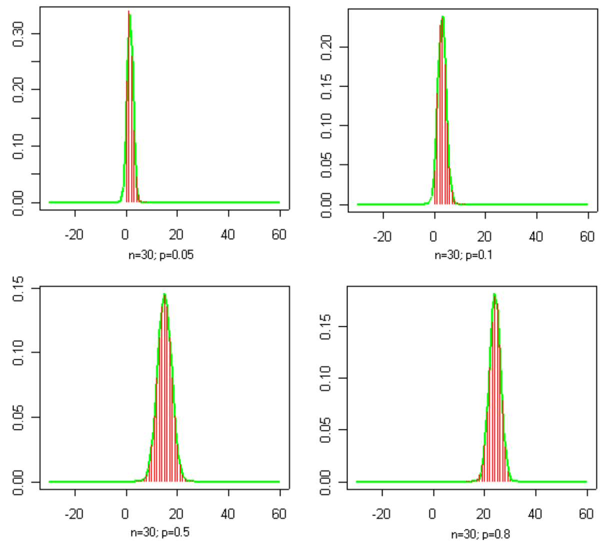
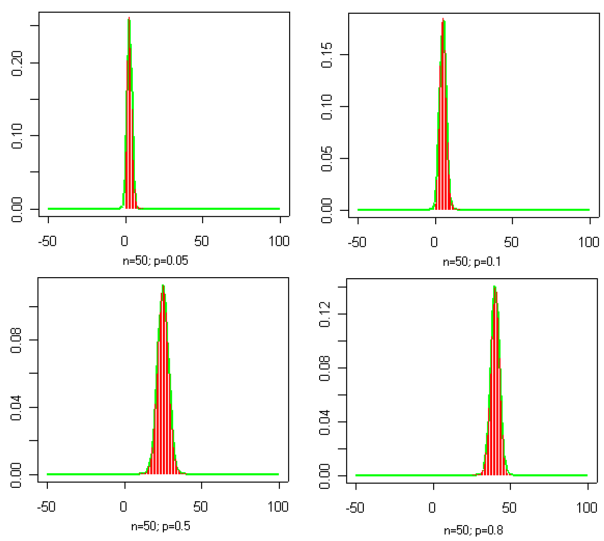


Figura 7.4: Ilustração da aproximação da dist. Binomial pela dist. Normal para $n = 50$



Exemplo 7.3 *Suponha que $X \sim P\,(230)$ e que queremos calcular $P\,(X = 241)$.*

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

,,