

2. INTRODUÇÃO À TEORIA DA PROBABILIDADE

Introdução. Noção de experiência aleatória

A teoria da Probabilidade tem como objectivo formular modelos de fenómenos naturais em que se supõe intervir o acaso, i.e., em que a partir do passado se não pode prever deterministicamente o futuro, mas para os quais se podem encontrar, em certas condições, taxas de realização constante, que poderão permitir certas previsões de índole geral.

Tais fenómenos dizem-se **fenómenos aleatórios**, i.e., são fenómenos sujeitos à influência do acaso e, como tal, fora do alcance do observador.

Definição 2.1

Considere-se uma experiência que verifica as seguintes características:

- pode repetir-se um grande número de vezes nas mesmas condições ou pelo menos em condições semelhantes;
- a sua realização dá um resultado de entre um conjunto de resultados possíveis w_1, w_2, \dots, w_N ;
- cada um dos resultados da experiência é imprevisível mas é possível considerar “estabilidade na frequência da sua ocorrência”.

Uma experiência com estas características diz-se ser uma **experiência aleatória**.

Exemplos de experiências aleatórias:

1. lançamento de um dado e registo do número de pontos que sai;
2. lançamento de uma moeda e observação da face que fica voltada para cima;
3. lançamento de dois dados ;
4. tempo de vida de uma pessoa, em anos;
5. tempo de trabalho de uma máquina até à primeira avaria.

Em cada um dos exemplos dados não é possível saber à priori o resultado que se irá obter. Os fenómenos aleatórios são caracterizados pela sua **imprevisibilidade** (fenómeno não determinístico) e pela sua **regularidade estatística** (observando-se o fenómeno um grande número de vezes, nas mesmas condições, a frequência relativa de cada resultado

possível do fenómeno tende a estabilizar, aproximando-se dum valor constante). Estas características foram já referidas na definição de experiência aleatória, dada acima.

Sendo assim, num fenómeno aleatório não se pode prever o resultado da próxima prova, mas pode fazer-se uma **previsão do resultado em média**.

Espaço de resultados. Noção de acontecimento

Definição 2.2

Chama-se **espaço de resultados** ou **espaço amostra** e representa-se por Ω ao conjunto de todos os resultados possíveis associados a uma experiência aleatória.

Para cada um dos exemplos citados acima temos os seguintes espaços de resultados:

1. $\Omega = \{1, 2, 3, 4, 5, 6\}$;
2. $\Omega = \{ \text{'valor'}, \text{'país'} \} = \{ \text{'V'}, \text{'P'} \} = \{1, 0\}$;
3. $\Omega = \{(1, 1), (1, 2), (1, 3), \dots, (6, 5), (6, 6)\}$;
4. $\Omega = \mathbb{N}$;
5. $\Omega = \mathbb{R}^+$.

Definição 2.3

Chama-se **acontecimento aleatório** a qualquer subconjunto do espaço de resultados.

Por exemplo

- no lançamento do dado, o acontecimento **A** – **saída de face par** pode representar-se por $\mathbf{A} = \{2, 4, 6\}$;
- na observação do tempo de trabalho de uma máquina até à primeira avaria, o acontecimento **B** – **duração entre 10 e 12 anos** é $\mathbf{B} = \{x : 10 < x < 12\}$.

Se um acontecimento é constituído por um único elemento diz-se **acontecimento elementar**, um acontecimento que não contém nenhum elemento diz-se **acontecimento impossível** e ao espaço Ω chama-se **acontecimento certo**.

Diz-se que um acontecimento se realiza sempre que o resultado de uma experiência é um elemento que pertence ao acontecimento.

Do que ficou dito verifica-se que há equivalência entre a noção de acontecimento e a noção de conjunto. Tem-se então um paralelismo entre álgebra de conjuntos e álgebra de acontecimentos. Consideremos as principais noções da

Álgebra dos Acontecimentos

1. Diz-se que A é **subacontecimento** de B e escreve-se $A \subset B$, se e só se a realização de A implica a realização de B ;
2. Dado um acontecimento A , chama-se **acontecimento complementar** ou **contrário** a A e representa-se por A^c ou \bar{A} , ao conjunto de todos os elementos de Ω que não estão em A .
3. Dados os acontecimentos A e B chama-se **união** de A com B e representa-se por $A \cup B$ ao acontecimento que consiste na realização de pelo menos um deles;
4. **produto** ou **intersecção** é o acontecimento AB ou $A \cap B$, que se realiza apenas quando ambos os acontecimentos se realizam;

Os acontecimentos A e B dizem-se **mutuamente exclusivos** ou **incompatíveis** se e só se a realização de um implica a não realização do outro, i.e., se e só se $A \cap B = \emptyset$, ou seja a intersecção é o acontecimento impossível;

5. Chama-se **diferença** dos acontecimentos A e B ao acontecimento $A - B = A \cap B^c$, i.e., ao acontecimento que se realiza se e só se A se realiza sem que B se realize.

Se $B \subset A$, $A - B$ é o acontecimento complementar de A em relação a B .

As propriedades estudadas na álgebra dos conjuntos (associatividade, comutatividade, idempotência, absorção, distributividade, leis de Morgan, dupla negação, complementaridade, citando as mais importantes), são válidas para acontecimentos.

Probabilidade de um acontecimento

Intuitivamente, a noção de probabilidade de um acontecimento é uma medida da possibilidade de ocorrência do acontecimento quando se realiza a experiência aleatória à qual o acontecimento está ligado.

A primeira definição de probabilidade conhecida, foi a sintetizada por Laplace no princípio do séc. XIX, sob a hipótese de **casos igualmente prováveis ou possíveis**, ou o chamado **princípio da simetria**.

A definição de Laplace dizia o seguinte:

- A probabilidade de realização de um dado acontecimento é igual ao quociente entre o número de casos favoráveis à realização desse acontecimento e o número total de casos possíveis, desde que todos os casos sejam igualmente prováveis.

Seja A o acontecimento “saída de face par” quando do lançamento de um dado equilibrado. Como há 3 casos favoráveis em 6 casos possíveis tem-se

$$P(A) = \frac{3}{6}.$$

A definição clássica de Laplace manteve-se até ao começo deste século, quando começaram a surgir críticas por ela apresentar diversos inconvenientes. Não era uma definição suficientemente geral pois fazia depender o cálculo das probabilidades do facto de os diferentes casos serem igualmente prováveis e numeráveis.

A regularidade estatística dos fenómenos aleatórios faz surgir uma outra teoria a **Teoria frequentista da probabilidade**, por analogia com a noção empírica de frequência. Surgiu no início do século e segundo ela a probabilidade de um acontecimento pode ser determinada observando a frequência relativa de ocorrência desse acontecimento numa sucessão numerosa de experiências aleatórias.

Efectuando n repetições duma experiência aleatória, seja n_A o número de vezes que se verificou o acontecimento A nessas n repetições. Devido ao princípio da regularidade estatística é de esperar que as frequências relativas $f_n(A) = n_A/n$ do acontecimento A numa sucessão de provas com um grande número de repetições sejam aproximadamente iguais a um número, digamos P ($0 \leq P \leq 1$).

A probabilidade é então interpretada como frequência limite, i.e., quando n grande tem-se $f_n(A) \simeq P(A)$.

Esta teoria considera como uma medição física (frequência relativa) um conceito teórico (probabilidade). A probabilidade P aparece como um objecto matemático, satisfazendo certas propriedades imediatas que resultam da definição de Laplace e da noção de frequência relativa.

No início do séc XX começou a sentir-se a necessidade de uma axiomatização da teoria das probabilidades, que permitisse ultrapassar a ambiguidade da certos conceitos e interpretações, mas partindo da observação da realidade e que a ela se aplicasse. A definição axiomática de probabilidade que iremos apresentar foi introduzida por Kolmogoroff.

Definição 2.4

Dada uma experiência aleatória, seja Ω o espaço de resultados associado. Chama-se **probabilidade P** , a uma aplicação que a cada acontecimento de Ω associa um número real satisfazendo o seguinte conjunto de axiomas:

$$\mathbf{A1)} \quad P(A) \geq 0 \quad \forall A \subset \Omega;$$

$$\mathbf{A2)} \quad P(\Omega) = 1;$$

$$\mathbf{A3)} \quad P(A \cup B) = P(A) + P(B) \quad \text{se} \quad A \cap B = \emptyset. \quad (\text{Axioma das probabilidades totais}).$$

Os axiomas apresentados referem-se ao caso de Ω ser finito. Quando Ω é infinito, o conjunto de axiomas está incompleto. Terá então que se considerar a generalização do axioma **A3)** ao caso de uma sucessão infinita de acontecimentos. Teremos então o axioma

$$\mathbf{A3^*)} \quad P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \quad \text{se} \quad A_i \cap A_j = \emptyset, \quad i \neq j \quad (\text{Axioma completo das probabilidades totais}).$$

Leis básicas das probabilidades

Muitas propriedades interessantes podem ser deduzidas daquele conjunto de axiomas. Vejamos algumas:

1. $P(A^c) = 1 - P(A)$.

Dem: Como $A \cup A^c = \Omega$ e $A \cap A^c = \emptyset$, tem-se $P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$, por A3), donde $P(A) + P(A^c) = 1 \Rightarrow P(A^c) = 1 - P(A)$

2. $P(\emptyset) = 0$; basta ter em conta que $\emptyset = \Omega^c$.

3. $A \subset B \Rightarrow P(A) \leq P(B)$.

Dem : Se $A \subset B \Leftrightarrow B = A \cup (B - A)$ e $A \cap (B - A) = \emptyset$, então $P(B) = P(A) + P(B - A) \Rightarrow P(B) \geq P(A)$, pois $P(B - A) \geq 0$, por A1).

4. $P(A) \leq 1$, é consequência imediata da propriedade anterior tendo em conta que $A \subset \Omega$.

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

Dem: $A = (A \cap B) \cup (A - B)$, onde $A \cap B$ e $A - B$ são disjuntos. Logo $P(A) = P(A \cap B) + P(A - B)$, donde se conclui $P(A - B) = P(A) - P(A \cap B)$.

6. Se $B \subset A \Rightarrow P(A - B) = P(A) - P(B)$, é um caso particular de 5).

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

Dem: $A \cup B = (A - B) \cup (A \cap B) \cup (B - A)$, todos eles disjuntos 2 a 2, logo $P(A \cup B) = P(A - B) + P(A \cap B) + P(B - A)$ e então por 5. vem $P(A \cup B) = P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) = P(A) + P(B) - P(A \cap B)$.

8. Generalização deste resultado:

Sendo A_1, A_2, \dots, A_n acontecimentos quaisquer

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - \dots - P(A_{n-1} \cap A_n) + P(A_1 \cap A_2 \cap A_3) + \dots + P(A_{n-2} \cap A_{n-1} \cap A_n) + \dots + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n).$$

A demonstração é feita por indução e deixa-se como exercício.

Como **corolário** deste resultado tem-se

9. Sendo A_1, \dots, A_n acontecimentos mutuamente exclusivos

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

Probabilidade condicional. Independência.

Consideremos a seguinte experiência aleatória:

Numa população finita, em que todos os indivíduos têm a mesma probabilidade de serem seleccionados, um investigador escolhe ao acaso um indivíduo. Uma vez que estamos a considerar casos equiprováveis, se n for a dimensão da população, a probabilidade de um elemento $\omega \in \Omega$ ser seleccionado é $P(\{\omega\}) = 1/n$. Sejam a e b duas características da população e A e B os acontecimentos “possuir a característica a ” e “possuir a característica b ”.

Suponhamos que um indivíduo ω escolhido ao acaso, possui a característica b , i.e., $\omega \in B$; pretende-se saber qual a probabilidade de ω pertencer a A .

Estamos num caso em que se tem como espaço de resultados o acontecimento B e se pretende saber a probabilidade de A se realizar sabendo que B se realizou.

Duas coisas podem acontecer:

- A realiza-se, i.e., $\omega \in (A \cap B)$;
- A não se realiza, i.e., $\omega \notin A$, ou seja, $\omega \in B - A$.

Então a probabilidade de $\omega \in A$ sob a condição de $\omega \in B$ é dada como o quociente entre o número de casos favoráveis a A e a B (que designaremos por $f(A \cap B)$) e o número de casos favoráveis a B ($f(B)$), ou seja

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

Definição 2.5

Chama-se **probabilidade condicional de A dado B** ou **probabilidade de A se B** e representa-se por $P(A|B)$, com $P(B) > 0$, a

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(AB)}{P(B)}. \quad (2.1)$$

Teorema 2.1 – Teorema das probabilidades compostas

Se $P(A) > 0$, $P(B) > 0$, tem-se

$$P(AB) = P(A) P(B|A) = P(B) P(A|B). \quad (2.2)$$

A demonstração é imediata a partir da definição 2.5.

Exercício 2.2

Obter a generalização a três acontecimentos, A , B e C tais que $P(A) > 0$, $P(B) > 0$ e $P(C) > 0$, i.e., mostrar que

$$\begin{aligned} P(ABC) &= P(A)P(B|A)P(C|AB) = P(B)P(C|B)P(A|BC) = \\ &= P(C)P(A|C)P(B|AC). \end{aligned}$$

Definição 2.6

Dois acontecimentos A e B dizem-se **independentes** se e só se

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

Note-se que esta definição é válida com $P(A) \geq 0$ e $P(B) \geq 0$.

Observações

- Da definição conclui-se que se A e B são independentes então $P(A|B) = P(A)$ se $P(B) > 0$ e $P(B|A) = P(B)$ se $P(A) > 0$.
- Independência não é equivalente a exclusividade mútua. Recorde-se que dois acontecimentos se dizem mutuamente exclusivos quando $A \cap B = \emptyset$.

Dados então A e B acontecimentos mutuamente exclusivos, tem-se $P(A \cap B) = 0$. Eles serão independentes se e só se $P(A) P(B) = 0$, i.e., $P(A) = 0$ ou $P(B) = 0$. Então, se para ambos os acontecimentos se verificar $P(A) > 0$ e $P(B) > 0$, sendo mutuamente exclusivos não podem ser independentes. Tem-se portanto:

Dados A e B acontecimentos mutuamente exclusivos, para os quais $P(A) > 0$ e $P(B) > 0 \Rightarrow A$ e B não independentes.

Ou, pensando no contra recíproco, **se A e B independentes e $P(A) > 0$ e $P(B) > 0 \Rightarrow A$ e B não mutuamente exclusivos.**

Como exemplo de experiências aleatórias que conduzem a acontecimentos independentes podemos considerar tiragens com reposição, lançamentos de um dado, de uma moeda, etc.

Teorema 2.2

Se A e B são independentes, também o são A e \overline{B} , \overline{A} e B e ainda \overline{A} e \overline{B} .

Dem: Vamos demonstrar o **1º caso** (A e \overline{B} independentes).

$$P(A \cap \overline{B}) = P[A - (A \cap B)] = P(A) - P(A \cap B)$$

Como A e B são independentes tem-se

$$= P(A) - P(A) P(B) = P(A)[1 - P(B)] = P(A) P(\overline{B}).$$

O 2º caso é análogo ao 1º, ficando como exercício.

Vejam os então o **3º caso** (\overline{A} e \overline{B}).

$$\begin{aligned} P(\overline{A} \cap \overline{B}) &= P(\overline{A \cup B}) = 1 - P(A \cup B) = \\ &= 1 - [P(A) + P(B) - P(A \cap B)] = 1 - [P(A) + P(B) - P(A) P(B)] = \\ &= 1 - P(A) - [P(B)(1 - P(A))] = P(\overline{A}) - P(B) P(\overline{A}) = \\ &= P(\overline{A})(1 - P(B)) = P(\overline{A}) P(\overline{B}). \end{aligned}$$

Extensão do conceito de independência

Dados n acontecimentos A_1, A_2, \dots, A_n dizem-se independentes se para qualquer subsucessão A_{i_1}, \dots, A_{i_k} daqueles acontecimentos, se tem

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}).$$

Como caso particular, os acontecimentos A, B e C dizem-se independentes se e só se

1. $P(ABC) = P(A) P(B) P(C)$;
2. $P(AB) = P(A)P(B)$;
3. $P(AC) = P(A)P(C)$;
4. $P(BC) = P(B)P(C)$.

Efectivamente, para garantir a independência de n acontecimentos, não é suficiente garantir a independência dois a dois. Vejamos dois exemplos, nos quais se consideram três acontecimentos e poderemos ver que as condições 2., 3. e 4. \nRightarrow 1. e 1. \nRightarrow 2., 3. e 4.

Exemplo 2.1

Suponhamos um tetraedro regular, com faces numeradas de 1 a 4. As faces estão ainda pintadas da seguinte forma: uma de verde, outra de amarelo, outra de vermelho e outra de verde + amarelo + vermelho. Vamos realizar a experiência aleatória que consiste no lançamento do tetraedro e observação da face em que ele se apoia. A probabilidade de cada face é a mesma e é igual a $1/4$.

Consideremos os acontecimentos:

A_1 – saída da cor vermelha;

A_2 – saída da cor amarela;

A3- saída da cor verde.

Então $P(A1) = P(A2) = P(A3) = 2/4 = 1/2$.

Serão estes três acontecimentos independentes?

$P(A1 \cap A2) = 1/4 = P(A1) P(A2)$ $P(A1 \cap A3) = 1/4 = P(A1) P(A3)$

$P(A2 \cap A3) = 1/4 = P(A2) P(A3)$.

Porém $P(A1 \cap A2 \cap A3) = 1/4 \neq 1/2 \times 1/2 \times 1/2 = P(A1).P(A2).P(A3)$.

Logo 2. 3. 4. \nrightarrow 1.

Exemplo 2.2

Consideremos A , B e C acontecimentos tais que:

$$P(A) = 0.60 \quad P(B) = 0.80 \quad P(C) = 0.50$$

$$P(A \cap B) = 0.48 \quad P(A \cap C) = 0.30 \quad P(B \cap C) = 0.38$$

Tem-se

$$P(A \cap B \cap C) = 0.24 = P(A).P(B).P(C)$$

$$P(A \cap B) = 0.48 = P(A).P(B)$$

$$P(A \cap C) = 0.30 = P(A).P(C)$$

$$P(B \cap C) = 0.38 \neq P(B).P(C).$$

Logo 1. \nrightarrow 2. 3. 4.

Teorema da probabilidade total. Teorema de Bayes.

Consideremos o seguinte problema:

Numa propriedade foram semeadas duas variedades de milho, A e B , sabendo-se que a quantidade de A é metade da de B . A frequência de uma maçaroca de milho roxo é de 6% para a variedade A e 3.6% para a variedade B .

1ª Questão: Pretende-se saber qual a probabilidade de que uma maçaroca apanhada ao acaso seja de milho roxo.

Ora, para termos uma maçaroca de milho roxo, ela pode provir de A ou de B . Seja R o acontecimento “sair milho roxo”. Sabe-se que $P(R|A) = 0.06$ e $P(R|B) = 0.036$. Ora o acontecimento R pode escrever-se como:

$R = (R \cap A) \cup (R \cap B)$, visto A e B constituírem uma partição, portanto

$$P(R) = P[(R \cap A) \cup (R \cap B)]$$

Então

$P(R) = P(R \cap A) + P(R \cap B)$ pois $R \cap A$ e $R \cap B$ são mutuamente exclusivos e $P(R) = P(A) P(R|A) + P(B) P(R|B) = 0.044$.

O problema que acabámos de resolver corresponde à situação formal exposta no

Teorema 2.3 – Teorema da probabilidade total

Sejam A_1, A_2, \dots, A_n acontecimentos definindo uma partição sobre Ω , i.e.,

$$A_1 \cup A_2 \cup \dots \cup A_n = \Omega \quad \text{e} \quad A_i \cap A_j = \emptyset \quad (i \neq j).$$

Se $P(A_i) > 0$, então para qualquer acontecimento $B \in \Omega$ tem-se

$$P(B) = \sum_{i=1}^n P(A_i) P(B|A_i). \quad (2.3)$$

Dem: Tem-se

$$B = B \cap \Omega = B \cap (\cup_{i=1}^n A_i) = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n).$$

Como os acontecimentos $B \cap A_i$ se excluem mutuamente pois são subconjuntos de A_i , i.e., $(B \cap A_i) \cap (B \cap A_j) = B \cap (A_i \cap A_j) = \emptyset, \forall i, j$ com $i \neq j$, tem-se então

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

2ª Questão: Para o mesmo problema suponhamos que é recolhida uma maçaroca de milho que se vê ser roxa. Qual a probabilidade de que ela seja da variedade A ?

Pretende-se então determinar $P(A|R)$.

Considerando a definição de probabilidade condicional temos

$$P(A|R) = \frac{P(A \cap R)}{P(R)} = \frac{P(A) \cdot P(R|A)}{P(A) \cdot P(R|A) + P(B) \cdot P(R|B)}.$$

A resolução desta questão corresponde formalmente ao

Teorema 2.4 – Teorema de Bayes

Sejam A_1, A_2, \dots, A_n acontecimentos formando uma partição de Ω , onde $P(A_i) > 0$. Seja B um outro acontecimento de Ω , tal que $P(B) > 0$. Então para $k = 1, \dots, n$ tem-se

$$P(A_k|B) = \frac{P(A_k) \cdot P(B|A_k)}{\sum_{i=1}^n P(A_i) \cdot P(B|A_i)}. \quad (2.4)$$

Dem: Por definição de probabilidade condicional tem-se $P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}$.

Relembrando que $P(A_i \cap B) = P(B|A_i) \cdot P(A_i)$ e usando o teorema da probabilidade total obtemos o resultado pretendido.

As probabilidades $P(A_i)$ são interpretadas como probabilidades à ‘priori’ que, na prática, são muitas vezes atribuídas pela pessoa que está a analisar o problema, que em muitos casos são subjectivas, traduzindo o grau de crença que ela tem na realização de cada A_i . Se no entanto temos a informação de que B se realizou, o valor de $P(A_i|B)$, calculado pela fórmula de Bayes, representa a probabilidade à ‘posteriori’.

Diz-se que este é um método de calcular a probabilidade da causa, se se conhece o efeito. Por isso o teorema de Bayes é também chamado **Teorema das causas**.

Variáveis aleatórias

Vimos já que uma experiência aleatória era um procedimento que nos levava à obtenção de um ou vários resultados sujeitos ao acaso. Algumas vezes os elementos de Ω são números reais: cumprimentos, produções, contagens, etc, outras vezes porém, Ω não é um conjunto numérico e podemos não estar interessados nos detalhes dos elementos que o constituem, mas sim numa descrição numérica associada à experiência. Consideremos por exemplo o espaço de resultados associado ao lançamento sucessivo de uma moeda, três vezes. O espaço de resultados é

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

Podemos estar interessados no número de faces que saem, passando a associar a cada elemento do espaço de resultados o número de faces observadas. Assim passamos a representar o espaço amostra por valores numéricos, fazendo corresponder a cada resultado da experiência os valores 0,1,2 ou 3. Estes valores podem ser olhados como valores assumidos por uma variável no decurso de uma experiência aleatória. A essa variável chama-se **variável aleatória**.

Definição 2.7

Chama-se **variável aleatória (v.a.)** e costuma representar-se por X , a uma função cujo valor é um número real determinado pelo resultado de uma experiência aleatória, i.e,

$$X : \Omega \rightarrow \mathbb{R}.$$

No exemplo referido a variável X toma o valor 0 quando se realiza o acontecimento $\{CCC\}$; toma o valor 1 quando se realiza o acontecimento $\{FCC, CFC, CCF\}$; toma o valor 2 quando se realiza $\{FCF, FFC, CFF\}$ e toma o valor 3 quando se realiza $\{FFF\}$.

A cada um dos valores de uma variável aleatória X podemos fazer corresponder uma probabilidade $P_X = P[X = x]$, definida como sendo a probabilidade do acontecimento que tem como imagem x por meio da aplicação X . Assim:

$$P[X = 0] = P(\{CCC\}) = 1/8; \quad P[X = 1] = P(\{FCC, CFC, CCF\}) = 3/8$$

$$P[X = 2] = P(\{FFC, CFF, FCF\}) = 3/8; \quad P[X = 3] = P(\{FFF\}) = 1/8.$$

Tipos de variáveis aleatórias

As variáveis aleatórias podem ser:

- **discretas** se assumem um conjunto finito ou infinito numerável de valores.

Exemplos:

- número de pintas que sai no lançamento de um dado;
- observação, a intervalos regulares, do número de pessoas em fila de espera na caixa de um supermercado;
- observação do sexo num conjunto de nascimentos.

- **contínuas** são as susceptíveis de tomar qualquer valor real pertencente a um intervalo dado. Este intervalo pode mesmo ser $(-\infty, +\infty)$, i.e., a recta real.

Exemplos:

- o peso de um indivíduo;
- o comprimento de um folha;
- qualquer unidade de medida.

Para uma dada experiência aleatória podemos estar interessados no estudo de uma única característica – **variável aleatória unidimensional** ou no estudo de um conjunto de k características – **variável aleatória multidimensional**, ou **vector aleatório**.

Uma conceito muito importante, associado a toda a variável aleatória é a sua

Função de distribuição cumulativa ou função de distribuição

Definição 2.8

Dada uma v. a. X , chama-se **função de distribuição cumulativa** da v. a. X e representa-se por $F(\cdot)$ ou $F_X(\cdot)$, à aplicação

$$F : \mathbb{R} \rightarrow [0, 1],$$

assim definida

$$F(x) = P[X \leq x]. \quad (2.5)$$

Exemplo 2.3

Cálculo da função de distribuição da v.a. X do exemplo estudado na página anterior.

$$F(x) = \begin{cases} 0 & x < 0; \\ 1/8 & 0 \leq x < 1; \\ 4/8 & 1 \leq x < 2; \\ 7/8 & 2 \leq x < 3; \\ 1 & x \geq 3. \end{cases}$$

Representação gráfica de $F(\cdot)$

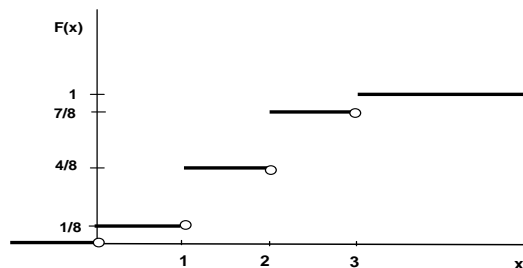


Figura 10: Gráfico da função distribuição cumulativa, F .

Trata-se de uma função em escada, onde os pontos de salto são os valores onde a v.a. está definida.

Para uma melhor caracterização desta função, vejamos, sem demonstração, as propriedades elementares.

Propriedades da função de distribuição cumulativa

1. $0 \leq F(x) \leq 1$

É consequência imediata da definição:

2. $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$; $F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$.

3. Dados dois números reais x_1 e x_2 tais que $x_1 < x_2$, tem-se

$$F(x_1) \leq F(x_2),$$

i.e., é uma função monótona não decrescente.

De facto tem-se

$$\begin{aligned} x_1 < x_2 &\Rightarrow] - \infty, x_1] \subset] - \infty, x_2] \Rightarrow \\ &\Rightarrow P(X \leq x_1) \leq P(X \leq x_2) \Rightarrow F(x_1) \leq F(x_2). \end{aligned}$$

4. $F(x)$ é contínua à direita, i.e., $\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$.

5. $P(X = a) = F(a) - F(a^-)$, onde $F(a^-) = \lim_{x \rightarrow a^-} F(x)$

Observação

O conhecimento da função de distribuição $F(\cdot)$ é equivalente ao conhecimento da lei de probabilidade P_X .

Efectivamente é trivial, a partir da definição, que o conhecimento de P_X implica o conhecimento de $F(x)$, pois $F(x) = P_X[X \leq x] = P[X \leq x]$.

Reciprocamente, mostraremos que o conhecimento de $F(x)$, arrasta o conhecimento de P_X , fazendo o cálculo da probabilidade dos vários tipos de intervalos.

- $P(X \leq x) = F(x)$ (por definição);
- $P(X < x) = P(X \leq x) - P(X = x) = F(x^-)$;
- $P(X \geq x) = 1 - P(X < x) = 1 - F(x^-)$;
- $P(X > x) = 1 - P(X \leq x) = 1 - F(x)$;
- $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$;
- $P(a < X < b) = P(X < b) - P(X \leq a) = F(b^-) - F(a)$;
- $P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F(b) - F(a^-)$;
- $P(a \leq X < b) = P(X < b) - P(X < a) = F(b^-) - F(a^-)$.

Definição 2.9

Duas variáveis aleatórias X e Y dizem-se **identicamente distribuídas** se têm iguais funções de distribuição, i.e., $F_X(x) = F_Y(y)$.

Exemplo 2.4

Consideremos as variáveis aleatórias X e Y definidas pelo lançamento de um dado do seguinte modo:

n ^o de face que sai	valor de X	valor de Y
1	1	2
2	2	0
3	1	2
4	0	1
5	2	1
6	2	2

Observe-se de facto que

$$\begin{array}{lll}
P(X = 0) = P(\{4\}) = 1/6 & \text{e} & P(Y = 0) = P(\{2\}) = 1/6 \\
P(X = 1) = P(\{1, 3\}) = 2/6 & \text{e} & P(Y = 1) = P(\{4, 5\}) = 2/6 \\
P(X = 2) = P(\{2, 5, 6\}) = 3/6 & \text{e} & P(Y = 2) = P(\{1, 3, 6\}) = 3/6
\end{array}$$

Logo as variáveis X e Y têm a mesma lei. Não são no entanto iguais, isto significaria que $X = Y$ sempre e isto não se verifica no nosso caso; por exemplo quando sai a face $\{1\}$ temos $X = 1$ e $Y = 2$; quando sai a face $\{4\}$ temos $X = 0$ e $Y = 1$.

Distribuições de probabilidade discretas

Como já dissémos, uma variável aleatória diz-se **discreta** se toma um número finito ou uma infinidade numerável de valores. Pode ainda definir-se à custa da sua função de distribuição: varia por saltos, sendo constante entre dois saltos consecutivos.

Consideremos n valores da v.a. X , x_1, \dots, x_n ; cada um destes valores ocorrendo com probabilidades p_1, \dots, p_n , respectivamente, i.e., $p_i = P[X = x_i]$. A esta função que associa a cada valor que a variável aleatória toma uma probabilidade chama-se **função massa de probabilidade**. Para uma v.a. X esta função satisfaz:

$$p_i \geq 0, i = 1, \dots, n \quad \sum_{i=1}^n p_i = 1.$$

Chama-se **distribuição de probabilidade** da v.a. X ao conjunto de pares (x_i, p_i) , i.e., aos valores da variável e respectivas probabilidades, que podemos dispor na forma:

$$X = \begin{Bmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{Bmatrix}$$

A função de distribuição cumulativa de uma variável aleatória discreta é, aplicando a definição, dada por

$$F(x) = P[X \leq x] = \sum_{x_i \leq x} P[X = x_i].$$

Exemplo 2.5

Consideremos a variável aleatória X com a seguinte distribuição de probabilidade:

$$X = \begin{Bmatrix} 1 & 2 & 3 \\ 1/6 & 1/2 & 2/6 \end{Bmatrix}$$

A função de distribuição desta variável é

$$F(x) = \begin{cases} 0 & \text{se } x < 1 \\ 1/6 & \text{se } 1 \leq x < 2 \\ 4/6 & \text{se } 2 \leq x < 3 \\ 1 & \text{se } x \geq 3, \end{cases}$$

Na figura seguinte apresenta-se a representação gráfica da distribuição de probabilidades de X e da respectiva função de distribuição cumulativa

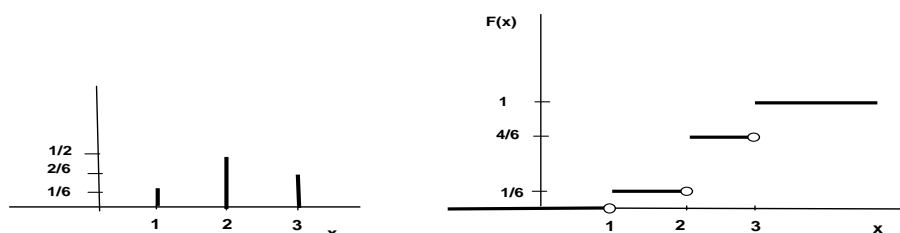


Figura 11: Gráfico da distribuição de probabilidades (à esquerda) e da função distribuição cumulativa(à direita).

Distribuições de probabilidade contínuas

Uma variável aleatória diz-se **contínua** se a sua função de distribuição cumulativa for contínua. De entre as variáveis aleatórias com função de distribuição contínua, só nos vão interessar as que são absolutamente contínuas, i.e, aquelas variáveis aleatórias para as quais existe uma função $f(\cdot)$ não negativa, definida em \mathbb{R} , excepto talvez num conjunto finito ou infinito numerável, tal que a função de distribuição $F(\cdot)$ verifica a relação:

$$F(x) = \int_{-\infty}^x f(t) dt \quad -\infty < x < \infty. \quad (2.6)$$

Sendo assim, a função de distribuição definida atrás é diferenciável, sendo

$$F'(x) = f(x)$$

(a menos de um número finito ou uma infinidade numerável de pontos).

Observações:

- Como $F(x) = \int_{-\infty}^x f(t) dt$, e $f(\cdot) \geq 0$ a função de distribuição representa a área da região compreendida entre $f(\cdot)$ e o eixo das abcissas, sendo o valor da área total =1, pois $F(+\infty) = 1$.

- Dados a e b , tais que $a < b$, tem-se

$$F(b) - F(a) = \int_{-\infty}^b f(t) dt - \int_{-\infty}^a f(t) dt = \int_a^b f(t) dt = P(a < X \leq b),$$

i.e., $F(b) - F(a)$ é a área compreendida entre $x = a$, $x = b$, o eixo das abcissas e a curva $f(\cdot)$.

Vejamos o significado dado à função f :

Consideremos um intervalo da forma $]x, x + \Delta x]$ ($\Delta x > 0$). Então

$$P(x < X \leq x + \Delta x) = F(x + \Delta x) - F(x),$$

que podemos interpretar como a quantidade de probabilidade no intervalo $]x, x + \Delta x]$. Sendo assim, o quociente

$$\frac{F(x + \Delta x) - F(x)}{\Delta x}$$

é a quantidade média de probabilidade naquele intervalo. Calculando

$$\lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}$$

se este limite existir, é igual a $F'(x)$ e representa a densidade de probabilidade no ponto x . E como já vimos, $F'(x) = f(x)$, convencionando escrever $f(x) = 0$ nos pontos em que $F'(\cdot)$ não existe.

Definição 2.10

A função $f(x)$ diz-se **função densidade de probabilidade (f.d.p.)** ou apenas **função densidade** se verificar as seguintes condições:

- $f(x) \geq 0 \quad \forall x$;
- $\int_{-\infty}^{+\infty} f(x) dx = 1$;

Da definição de função densidade e das propriedades da função de distribuição cumulativa temos a seguinte propriedade:

$$\int_a^b f(x) dx = F(b) - F(a) = P(a < X \leq b) = P(a \leq X \leq b).$$

Exemplo 2.6

Suponhamos que sobre um dado segmento de recta (a, b) , se escolhe um ponto ao acaso, i.e., tal que a probabilidade de escolha seja independente da posição. A densidade de probabilidade deve ser então considerada constante, i.e.,

$$f(x) = \begin{cases} c & a < x < b \\ 0 & x \leq a \text{ ou } x \geq b. \end{cases}$$

Para que $f(x)$ seja função densidade deve ser não negativa e verificar

$$\int_a^b f(x) dx = \int_a^b c dx = 1 \Rightarrow c = \frac{1}{b-a}$$

donde

$$f(x) = \begin{cases} 1/(b-a) & a < x < b \\ 0 & x \leq a \text{ ou } x \geq b \end{cases}$$

Cálculo da função de distribuição:

$$\text{se } x < a \quad F(x) = 0$$

$$\text{se } a \leq x < b \quad F(x) = \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a}$$

$$\text{se } x \geq b \quad F(x) = 1.$$

Sendo assim

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

Representação gráfica da função densidade e da função de distribuição:

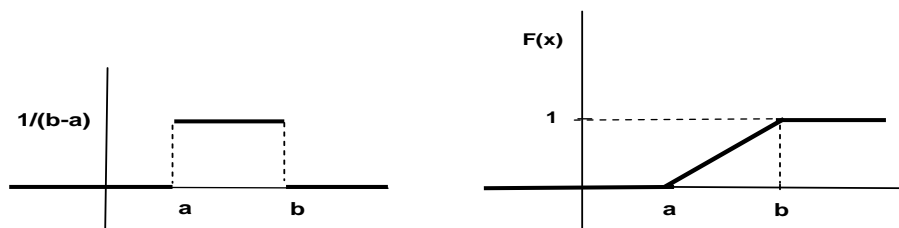


Figura 12: Gráfico da função densidade de probabilidades (à esquerda) e da função distribuição cumulativa (à direita).

A distribuição acabada de estudar chama-se **distribuição uniforme** no intervalo (a,b) .

Funções de variáveis aleatórias

Seja X uma variável aleatória e $Y = \varphi(X)$ uma função de X , logo é também uma variável aleatória, tendo portanto uma distribuição. Conhecida a função de distribuição $F(x)$ de X , pretende-se determinar $G(y)$ de Y .

1º caso

Se X é uma variável aleatória discreta, também $Y = \varphi(X)$ é discreta.

Assim, sendo $x_1, x_2, \dots, x_n, \dots$ os valores de X , com probabilidades $p_1, p_2, \dots, p_n, \dots$, respectivamente, $\varphi(X)$ terá os valores $\varphi(x_1), \varphi(x_2), \dots, \varphi(x_n), \dots$ com probabilidades $p_1, p_2, \dots, p_n, \dots$, respectivamente.

Exemplo 2.7

Seja X a v. a. com a seguinte distribuição de probabilidades

$$\begin{array}{ccc} 2 & 4 & 8 \\ 1/4 & 1/2 & 1/4 \end{array}$$

Se $Y = 2X$, então Y terá a seguinte distribuição de probabilidades

$$\begin{array}{ccc} 4 & 8 & 16 \\ 1/4 & 1/2 & 1/4 \end{array}$$

No caso de $\varphi(x_i) = \varphi(x_j)$ ter-se-á o valor da probabilidade $p_i + p_j$.

2º caso

Se X é uma v.a. contínua com função densidade $f_X(x)$ e $Y = \varphi(X)$, a variável transformada, com $\varphi(\cdot)$ função estritamente monótona (crescente ou decrescente) e com derivada em todos os pontos do respectivo domínio. Então a função densidade da v.a. Y , $f_Y(y)$, é assim definida

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|, \quad \text{com } x = \varphi^{-1}(y).$$

Dem: Consideremos $\varphi(\cdot)$ uma função decrescente. Então a função de distribuição de Y , $F_Y(y)$, vem

$$\begin{aligned} F_Y(y) &= P[Y \leq y] = P[\varphi(X) \leq y] = P[X \geq \varphi^{-1}(y)] = \\ &= 1 - P[X \leq \varphi^{-1}(y)] = 1 - F_X[\varphi^{-1}(y)]. \end{aligned}$$

Derivando $F_Y(y)$ em ordem a y e tendo em conta a regra da derivada da função composta, temos

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{d[1 - F_X[\varphi^{-1}(y)]]}{dy} = \frac{d}{dx}[1 - F_X(x)] \frac{dx}{dy} = -f_X(x) \frac{dx}{dy}.$$

Procedendo de modo análogo no caso de $\varphi(\cdot)$ ser uma função crescente, obtém-se

$$f_Y(y) = f_X(x) \frac{dx}{dy},$$

o que combinando os dois resultados nos permite escrever

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|, \quad \text{com } x = \varphi^{-1}(y).$$

Exemplo 2.8

Seja X a v. a. com densidade $f_X(x) = 2x$, se $0 < x < 1$, e 0 fora daquele intervalo. Considere-se a função $Y = 3X + 1$, então

$$f_Y(y) = 2x \times \frac{1}{3} \text{ com } x = \frac{y-1}{3}, \text{ portanto } f_Y(y) = 2(y-1)/9, \text{ para } 1 < y < 4.$$

Se a função $\varphi(\cdot)$ não for monótona, divide-se o seu domínio em intervalos de monotonia, aplicando-se o resultado anterior a cada subintervalo e somando as expressões obtidas.

Exemplo 2.9

Seja $Y = \varphi(X) = X^2$.

Tem-se $X = +\sqrt{Y}$, $0 \leq X < +\infty$ e $X = -\sqrt{Y}$, $-\infty < X \leq 0$.

Donde

$$f_Y(y) = f_X(\sqrt{y}) \left| \frac{1}{2\sqrt{y}} \right| + f_X(-\sqrt{y}) \left| -\frac{1}{2\sqrt{y}} \right| = \frac{1}{2\sqrt{y}} [f_X(\sqrt{y}) + f_X(-\sqrt{y})] \quad y \geq 0$$

Para $y < 0$ tem-se $f_Y(y) = 0$.

O resultado anterior pode ser derivado directamente a partir da função de distribuição:

$$F_Y(y) = P[Y \leq y] = P[X^2 \leq y] = P[-\sqrt{y} \leq X \leq \sqrt{y}] = F_X(\sqrt{y}) - F_X(-\sqrt{y}).$$

Por derivação tem-se

$$f_Y(y) = \frac{dF_Y(y)}{dy} = f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}}.$$

Exercício 2.3 – transformação uniformizante

Seja X uma v. a. contínua qualquer. Verificar que $Y = F(X)$, onde $F(x)$ é a função de distribuição de X , é uma variável aleatória com distribuição uniforme no intervalo $[0, 1]$.

Nota: Ver Teorema 11.

Vectores Aleatórios

Em muitas situações em que realizamos uma experiência aleatória não estamos interessados apenas num único valor, mas em registar um conjunto de características associadas a cada um dos resultados da experiência aleatória.

Por exemplo podemos pretender medir a quantidade de precipitado P e do volume V de gás numa experiência química; a altura de uma árvore e o diâmetro do tronco à altura do peito, etc.

Consideremos então brevemente o caso da distribuição conjunta de duas variáveis aleatórias, quer no caso discreto quer no caso contínuo. O caso de três ou mais variáveis é apenas uma generalização.

Definição 2.11

Chama-se par aleatório (X, Y) à aplicação

$$(X, Y) : \Omega \rightarrow \mathbb{R}^2.$$

Um par aleatório diz-se **discreto** se ambas as componentes do par forem variáveis aleatórias discretas; diz-se **contínuo** se ambas forem contínuas e diz-se **misto** se uma componente for contínua e outra discreta. Este último caso não será considerado nestes apontamentos.

Nota: Chama-se atenção para o facto de o conceito apresentado a seguir, função distribuição cumulativa conjunta, não ter sido tratado nas aulas deste ano lectivo. Entendemos, porém, que não deve ser retirado deste material de consulta, por ser um conceito importante para uma completa caracterização de um vector aleatório.

Definição 2.12

Dado o par aleatório (X, Y) , chama-se **função de distribuição cumulativa conjunta** ou apenas **função de distribuição conjunta** e representa-se por $F(., .)$ à função real de duas variáveis reais x e y

$$F : \mathbb{R}^2 \rightarrow [0, 1]$$

definida por

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

$F(x, y)$ goza de propriedades análogas às referidas atrás para a função de distribuição de uma v.a., i.e.,

1. $0 \leq F(x, y) \leq 1$
2. $F(-\infty, y) = F(x, -\infty) = 0$; $F(+\infty, +\infty) = 1$
3. Toda a função de distribuição, $F(x, y)$, é não decrescente relativamente a cada um dos argumentos.
4. Toda a função de distribuição, $F(x, y)$, é contínua à direita em relação a qualquer dos argumentos, i.e.,

$$\lim_{x \rightarrow x_0^+} F(x, y) = F(x_0, y); \quad \lim_{y \rightarrow y_0^+} F(x, y) = F(x, y_0)$$

5. Há porém uma propriedade que não tem correspondente no caso de uma variável aleatória, que é a seguinte:

$F(x, y)$ é função de distribuição se e só se verificar

$$F(x_1, y_1) - F(x_0, y_1) - F(x_1, y_0) + F(x_0, y_0) \geq 0 \quad \text{com} \quad x_0 < x_1, \quad y_0 < y_1.$$

Exemplo 2.10

Consideremos a seguinte função

$$F(x, y) = \begin{cases} 0 & x + y < 0 \\ 1 & x + y \geq 0 \end{cases}$$

Trata-se de uma função que verifica as propriedades 1, 2, 3 e 4 enunciadas acima. Porém, vejamos que a propriedade 5 não é verificada:

Tomando $x_0 = y_0 = -1$ e $x_1 = y_1 = 2$, tem-se

$$F(2, 2) - F(-1, 2) - F(2, -1) + F(-1, -1) = -1 < 0.$$

Portanto, $F(x, y)$ não é função de distribuição.

Tendo a distribuição conjunta de duas ou mais variáveis, podemos estar interessados em estudar o comportamento de cada uma delas separadamente, aquilo a que se chama **as margens**.

Chamam-se **funções de distribuição marginais** de X e Y e representam-se por $F_X(x)$ e $F_Y(y)$ a

$$F_X(x) = \lim_{y \rightarrow +\infty} F(x, y) = F(x, +\infty)$$

$$F_Y(y) = \lim_{x \rightarrow +\infty} F(x, y) = F(+\infty, y)$$

Começemos por estudar o **par aleatório discreto**

(X, Y) diz-se um par aleatório **discreto** se toma os valores (x_i, y_j) com probabilidades $p_{ij} = P[X = x_i, Y = y_j]$.

Chama-se **distribuição de probabilidades conjunta** do par (X, Y) ao conjunto de valores (x_i, y_j) e respectivas probabilidades p_{ij} , se e só se são verificadas as seguintes condições:

$$p_{ij} \geq 0 \quad \forall i, j \quad \text{e} \quad \sum_{i,j} p_{ij} = 1.$$

Os pares aleatórios discretos representam-se muitas vezes sob a forma de uma **tabela de contingência**, contendo os valores das variáveis e as probabilidades conjuntas:

	Y	y_1	y_2	...	y_n	
X						
x_1		p_{11}	p_{12}	...	p_{1n}	$p_{1.}$
x_2		p_{21}	p_{22}	...	p_{2n}	$p_{2.}$
.	
.	
.	
x_m		p_{m1}	p_{m2}	...	p_{mn}	$p_{m.}$
		$p_{.1}$	$p_{.2}$...	$p_{.n}$	1

A $p_{i.} = \sum_{j=1}^n p_{ij}$ e $p_{.j} = \sum_{i=1}^m p_{ij}$ chamam-se **probabilidades marginais** de X e Y respectivamente.

A **função de distribuição cumulativa** de (X, Y) é neste caso dada por

$$F(x, y) = P[X \leq x, Y \leq y] = \sum_{x_i \leq x} \sum_{y_j \leq y} P[X = x_i, Y = y_j]$$

As **funções de distribuição marginais** de X e Y , $F_X(x)$ e $F_Y(y)$ são no caso discreto dadas por

$$F_X(x) = \sum_{x_i \leq x} p_i.$$

$$F_Y(y) = \sum_{y_j \leq y} p_{.j}$$

Define-se **probabilidade condicional** de X dado $Y = y_j$ como

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{.j}}$$

e **probabilidade condicional** de Y dado $X = x_i$ como

$$P(Y = y_j | X = x_i) = \frac{p_{ij}}{p_i}.$$

Consideremos agora o **par aleatório contínuo**.

Definição 2.13

Um vector aleatório bidimensional (X, Y) diz-se **contínuo** (absolutamente contínuo) se a sua função de distribuição $F(x, y)$ é dada por

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) \, dudv$$

onde $f(., .) \geq 0$ é a **função densidade conjunta** do par aleatório (X, Y) .

Uma função $f(x, y)$ é **função densidade conjunta** do par aleatório (X, Y) se e só se verificar as seguintes condições:

- $f(x, y) \geq 0$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx dy = 1.$

Se B é um acontecimento de \mathbb{R}^2 , a probabilidade de se ter $(X, Y) \in B$ é assim obtida:

$$P[(X, Y) \in B] = \int \int_B f(x, y) \, dx dy.$$

Pela definição e pelas propriedades do integral vem que

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y).$$

Define-se **densidade marginal de X** como

$$f_X(x) = f(x, \cdot) = \int_{-\infty}^{+\infty} f(x, y) dy$$

e analogamente, a **densidade marginal de Y** é

$$f_Y(y) = f(\cdot, y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

A **densidade condicional de X dado $Y = y$** é definida por

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}. \quad f_Y(y) > 0$$

Analogamente para a **densidade condicional de Y dado $X = x$** .

Exemplo 2.11

Considere a seguinte função densidade:

$$f(x, y) = \begin{cases} kxy & \text{se } 1 < x < y < 2 \\ 0 & \text{o.v. de } (x, y) \end{cases}$$

1. Determine k .
2. Determine as funções densidade marginais de X e Y .
3. Determine a densidade condicional $f(y|x)$.

Resolução:

1. $f(x, y)$ é uma função densidade se $kxy \geq 0 \Rightarrow k \geq 0$ e $\int_{\mathbb{R}^2} f(x, y) dx dy = 1$, i.e.,

$$\int_1^2 dx \int_x^2 kxy dy = 1 \Rightarrow k \left[x^2 - \frac{x^4}{8} \right]_1^2 = 1 \Rightarrow k \frac{9}{8} = 1 \Rightarrow k = \frac{8}{9}.$$

2. Densidade marginal de X

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \int_x^2 kxy dy = 16x/9 - 4x^3/9 \quad \text{se } 1 < x < 2.$$

- 3.

$$f(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{8xy/9}{16x/9 - 4x^3/9} = \frac{2y}{4 - x^2} \quad \text{se } x < y < 2.$$

Exercício 2.4

Para a função dada no exemplo anterior calcule $F_X(x)$, $f_Y(y)$ e $f(x|y)$.

Independência de variáveis aleatórias

Definição 2.14

Dado o par aleatório (X, Y) , as variáveis X e Y dizem-se **independentes** se e só se

$$F(x, y) = F_X(x)F_Y(y) \quad \forall (x, y) \in \mathbb{R}^2$$

A definição de independência pode ser dada de modo a permitir uma utilização mais fácil, distinguindo o caso discreto e o caso contínuo,

No caso de (X, Y) ser um **par aleatório discreto** as variáveis X e Y dizem-se **independentes** se e só se $p_{ij} = p_{i.} p_{.j} \quad \forall i, j$.

Se (X, Y) é um **par aleatório contínuo**, as variáveis X e Y dizem-se **independentes** se e só se $f(x, y) = f_X(x) f_Y(y) \quad \forall (x, y) \in \mathbb{R}^2$.

Exercício 2.5

Seja (X, Y) um par aleatório com função densidade conjunta assim definida:

$$f(x, y) = \exp(-x - y) \quad \text{para } x \geq 0, y \geq 0.$$

Verifique que X e Y são variáveis aleatórias independentes.

Características numéricas de uma variável aleatória e de um par aleatório

A uma variável aleatória quer no caso unidimensional, quer multidimensional podemos associar características numéricas (parâmetros) que nos dão informação sobre a variável.

Valor Médio

Definição 2.15

Dada uma v.a. X chama-se **valor médio, esperança matemática, valor esperado** ou **média** e representa-se por $E[X]$, μ_X ou simplesmente μ a quantidade assim definida

$$E[X] = \sum_{i=1}^n x_i p_i \quad \text{se } X \text{ v.a. discreta com distribuição } (x_i, p_i) \quad i = 1, \dots, n$$

ou

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx \quad \text{se } X \text{ v.a. contínua com densidade } f(x).$$

Observação: Se X for v.a. discreta com uma infinidade numerável de valores tem-se $E[X] = \sum_{i=1}^{\infty} x_i p_i$. Neste caso só existe valor médio se a série ¹ for absolutamente convergente.

Analogamente, no caso contínuo, só existe valor médio se o integral for absolutamente convergente.

Se X é uma v.a. e $Y = \varphi(X)$ é uma função real de variável real, tem-se

$$E[\varphi(X)] = \sum_i \varphi(x_i) p_i \quad \text{se } X \text{ é v.a. discreta com distribuição } (x_i, p_i);$$

$$E[\varphi(X)] = \int_{-\infty}^{+\infty} \varphi(x) f(x) dx \quad \text{se } X \text{ é v.a. contínua com f.d.p. } f(x).$$

Mais uma vez, para que exista valor médio exige-se a convergência absoluta da série (no caso de se tratar de uma v.a. discreta com uma infinidade de valores) ou a convergência absoluta do integral.

¹A noção de série e claro, todos os conceitos a ela associados, não integraram o programa da disciplina Matemática e Informática no ano lectivo passado. Como consideramos que se trata de um conceito de extrema importância em Probabilidade e Estatística optámos por o incluir nos apontamentos, sempre que necessário, não constituindo porém assunto para avaliação. Para que os alunos consigam compreender do que estamos a falar deixamos a ideia do que é uma série (numérica): dada uma sucessão $u_n, n \in \mathbb{N}$, chama-se série a $u_1 + u_2 + u_3 + \dots + u_n + \dots$, i.e., à expressão soma de todos os termos da sucessão e costuma representar-se por $\sum_{n=1}^{\infty} u_n$. Se aquela soma for um número real, diz-se que a série é convergente.

Exemplo 2.12

Seja X uma v.a. que representa a vida em horas de um certo tipo de tubo, cuja f.d.p. é dada por

$$f(x) = \frac{20000}{x^3} \quad x > 100; \quad = 0 \quad \text{outros valores.}$$

Qual será o tempo de vida esperado para este tipo de tubos?

A resposta é dada pelo valor médio da v.a. X .

$$E[X] = \int_{100}^{+\infty} x \cdot 20000/x^3 \, dx = 200.$$

O tempo de vida esperado é portanto de 200 horas.

Sendo (X, Y) um par aleatório e $g(X, Y)$ uma função real de (X, Y) . Define-se

$$E[g(X, Y)] = \sum_i \sum_j g(x_i, y_j) p_{ij}, \quad \text{no caso discreto}$$

$$E[g(X, Y)] = \int \int_{R^2} g(x, y) f(x, y) \, dx dy, \quad \text{no caso contínuo.}$$

De novo se exige aqui a convergência absoluta da série dupla e do integral.

Propriedades do valor médio

1. Linearidade

- $E[a] = a$.

Nota: Observe-se que, dizer que $X = a$, significa que a variável aleatória toma unicamente o valor a , i. e., a sua função de distribuição é da forma

$$F(x) = \begin{cases} 0 & x < a; \\ 1 & x \geq a. \end{cases}$$

Uma variável aleatória tal como esta diz-se *degenerada* (de facto toda a probabilidade está concentrada no ponto a).

- $E[a + bX] = a + b E[X]$.

Dem: Suponhamos X uma v.a. discreta com distribuição de probabilidade (x_i, p_i) , ficando como exercício o caso contínuo:

$$E[a + bX] = \sum_i (a + b x_i) p_i = a \sum_i p_i + b \sum_i x_i p_i = a + b E[X].$$

- $E[\varphi(X) + \psi(X)] = E[\varphi(X)] + E[\psi(X)]$

Dem:

$$E[\varphi(X) + \psi(X)] = \int_R (\varphi(x) + \psi(x))f(x) dx = \int_R \varphi(x) f(x) dx + \int_R \psi(x) f(x) dx = E[\varphi(X)] + E[\psi(X)]$$

Foi considerada aqui a demonstração no caso contínuo, ficando como exercício o caso discreto.

- 2. Positividade** Se $X \geq 0$, i.e. a variável toma apenas valores ≥ 0 , tem-se $E[X] \geq 0$, como é imediato verificar.

- 3. $\inf(X) \leq E[X] \leq \sup(X)$**

É imediato, bastando ter em conta, no caso discreto por exemplo, que $\inf(x_i) \leq x_i \leq \sup(x_i)$.

- 4. Aditividade** $E[X \pm Y] = E[X] \pm E[Y]$

Dem: Vejamos no caso discreto

$$E[X+Y] = \sum_{i,j} (x_i+y_j)p_{ij} = \sum_{i,j} x_i p_{ij} + \sum_{i,j} y_j p_{ij} = \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij} = \sum_i x_i p_{i.} + \sum_j y_j p_{.j} = E[X] + E[Y].$$

- 5.** Se X e Y são variáveis aleatórias independentes, tem-se

$$E[XY] = E[X]E[Y]$$

Dem: Consideremos agora o caso contínuo

Se X e Y são variáveis aleatórias independentes com funções densidade $f_X(x)$ e $f_Y(y)$, respectivamente, tem-se $f(x, y) = f_X(x)f_Y(y)$. Tendo em conta isto obtem-se facilmente

$$E[XY] = \int \int_{R^2} xy f(x, y) dx dy = \int x f(x) dx \int y f(y) dy = E[X] E[Y].$$

Nota: O recíproco não é verdadeiro:

Vejamos um contra-exemplo:

Consideremos X e Y duas variáveis aleatórias com a seguinte distribuição de probabilidades

	Y	-1	0	1
X				
0		0	$1/3$	0
1		$1/3$	0	$1/3$

Tem-se $E[X] = 2/3$ e $E[Y] = 0$ e ainda $E[XY] = 0$. Porém verifica-se imediatamente que X e Y não são independentes, por exemplo tem-se $p_{11} = 0$ e $p_{1.} = 1/3$ e $p_{.1} = 1/3$, logo $p_{11} \neq p_{1.} p_{.1}$.

Donde $E[XY] = E[X] E[Y]$ e porém X e Y não são independentes.

6. Desigualdade de Schwarz

Se $E[X^2]$ e $E[Y^2]$ existem então $E^2[XY] \leq E[X^2]E[Y^2]$.

Dem:

Seja $Z = X + \lambda Y$ uma v.a. onde λ é um parâmetro real.

Calculando $E[Z^2] = E[X^2] + 2\lambda E[X.Y] + \lambda^2 E[Y^2] \geq 0$, pela propriedade 2.

Esta desigualdade é verificada para qualquer valor de λ , se o binómio discriminante

$$4 E^2[X.Y] - 4E[X^2]E[Y^2] \leq 0, \text{ donde se tem } E^2[X.Y] \leq E[X^2]E[Y^2].$$

Como **corolário** desta propriedade tem-se:

$$E^2[X] \leq E[X^2]$$

Este resultado mostra-nos que se $E[X^2]$ existe (o integral ou a série são convergentes), existe também $E[X]$.

7. $|E[\varphi(X)]| \leq E[|\varphi(X)|]$.

Deixa-se ficar como exercício a demonstração desta propriedade.

Vejamos o seguinte teorema, que enunciamos sem demonstração e nos dá a relação existente entre funções de variáveis aleatórias independentes.

Teorema 2.5

Dadas duas variáveis aleatórias independentes, (X, Y) , sejam $\varphi(\cdot)$ e $\phi(\cdot)$, duas funções injectivas, definindo duas outras variáveis aleatórias, $U = \varphi(X)$ e $V = \phi(Y)$. As variáveis aleatórias U e V são também independentes.

Variância

O valor médio é uma medida do centro de uma distribuição. Interessa, porém, considerar uma medida da dispersão dessa distribuição. Definindo o **desvio** de uma v.a. X relativamente ao seu valor médio μ , como $X - \mu$, tem-se, usando propriedades do valor médio,

$$E[X - \mu] = 0.$$

A medida da dispersão deverá então ter em conta a grandeza dos desvios e não o seu sinal. Essa medida é a **variância**, que se representa por $\mathbf{Var}[X]$, σ_X^2 ou simplesmente σ^2 e é assim definida:

$$\mathbf{Var}[X] = E[(X - \mu)^2] = E[X^2] - \mu^2. \quad (2.7)$$

$\sigma_X = \sqrt{\mathbf{Var}[X]}$ chama-se **desvio padrão**.

Propriedades da variância

1. $\mathbf{Var}[X] \geq 0$, o que é imediato pela própria definição. A igualdade verifica-se apenas no caso discreto, sendo X a v.a. que toma apenas um único valor; a esta variável chama-se **v.a. degenerada**; no caso contínuo tem-se sempre $\mathbf{Var}[X] > 0$.

2. $\mathbf{Var}[a + bX] = b^2 \mathbf{Var}[X]$.

$$\begin{aligned} \text{Dem: } \mathbf{Var}[a + bX] &= E[((a + bX) - (a + b\mu))^2] = E[(bX - b\mu)^2] \\ &= b^2 E[(X - \mu)^2] = b^2 \mathbf{Var}[X]. \end{aligned}$$

Para o desvio padrão temos a propriedade sob a forma

$$\sigma_{(a+bX)} = |b| \sigma_X$$

3. Se X e Y são variáveis aleatórias independentes tem-se

$$\mathbf{Var}[X \pm Y] = \mathbf{Var}[X] + \mathbf{Var}[Y]$$

Dem: Vejamos a demonstração no caso da diferença, ficando como exercício o caso da soma.

$$\begin{aligned} \mathbf{Var}[X - Y] &= E[((X - Y) - (\mu_X - \mu_Y))^2] = E[((X - \mu_X) - (Y - \mu_Y))^2] = \\ &= E[(X - \mu_X)^2 + (Y - \mu_Y)^2 - 2(X - \mu_X)(Y - \mu_Y)] = \mathbf{Var}[X] + \mathbf{Var}[Y] - \\ &- 2 E[(X - \mu_X)(Y - \mu_Y)]. \end{aligned}$$

Vamos então provar que, sendo X e Y independentes se tem

$$E[(X - \mu_X)(Y - \mu_Y)] = 0.$$

De facto após breves cálculos obtém-se

$$E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y = 0 \text{ pela propriedade 5., do valor médio.}$$

A expressão $E[(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{Y} - \boldsymbol{\mu}_Y)]$ mede a variação conjunta das duas variáveis X e Y e designa-se por **covariância**. Representa-se por $\mathbf{Cov}(\mathbf{X}, \mathbf{Y})$ ou $\boldsymbol{\sigma}_{\mathbf{X}, \mathbf{Y}}$.

Atendendo aos cálculos apresentados pode escrever-se

$$\mathbf{Cov}[\mathbf{X}, \mathbf{Y}] = \mathbf{E}[\mathbf{XY}] - \mathbf{E}[\mathbf{X}]\mathbf{E}[\mathbf{Y}]. \quad (2.8)$$

Propriedades da covariância

1. Se X e Y são variáveis aleatórias independentes tem-se $Cov[X, Y] = 0$.

A demonstração desta propriedade é imediata tendo em conta a expressão (2.9).

Nota: O recíproco da propriedade não é verdadeiro.

2. $Var[X \pm Y] = Var[X] + Var[Y] \pm 2 Cov[X, Y]$

o que é também imediato tendo em conta a demonstração da propriedade 3. da variância.

Exercício 2.6 Generalização desta propriedade:

Sejam X_1, X_2, \dots, X_n n variáveis aleatórias, tem-se então

$$Var\left[\sum_i X_i\right] = \sum_i Var[X_i] + 2 \sum_{i < j} Cov[X_i, X_j]. \quad (2.9)$$

Para as variáveis do exercício anterior, chama-se **matriz de variância-covariância** à matriz quadrada de ordem n , $\Sigma = [\sigma_{ij}]$ onde $\sigma_{ij} = Cov[X_i, X_j]$ e $\sigma_{ii} = Var[X_i]$.

3. $Cov[a + bX, c + dY] = bd Cov[X, Y]$, com $b \neq 0$ e $d \neq 0$

$$\begin{aligned} \text{Dem: } Cov[a + bX, c + dY] &= E\{[(a + bX) - (a + b\mu_X)][(c + dY) - (c + d\mu_Y)]\} = \\ &= E\{[bX - b\mu_X][dY - d\mu_Y]\} = bd E[(X - \mu_X)(Y - \mu_Y)] = bd Cov[X, Y]. \end{aligned}$$

4. $|Cov[X, Y]| \leq \sigma_X \sigma_Y$.

é uma consequência imediata da desigualdade de Schwarz, aplicando-a a $(X - E[X])$ e $(Y - E[Y])$.

Como consequência da propriedade 3. vemos que a covariância depende das unidades em que se exprimem as variáveis aleatórias X e Y . Sendo assim, é importante a

introdução de um parâmetro para caracterizar a intensidade da ligação entre X e Y , que não dependa das unidades.

Temos assim o **coeficiente de correlação** que é definido como:

$$\rho = \rho_{X,Y} = \frac{Cov[X, Y]}{\sigma_X \sigma_Y}. \quad (2.10)$$

desde que $\sigma_X > 0$ e $\sigma_Y > 0$.

Propriedades do coeficiente de correlação

1. $-1 \leq \rho_{X,Y} \leq 1$

É uma consequência imediata da propriedade 4. da covariância.

2. Se X e Y são v. a. independentes tem-se $\rho_{X,Y} = 0$.

É imediato da definição de coeficiente de correlação () e da propriedade 1. da covariância.

3. O coeficiente de correlação não se altera quando as variáveis sofrem uma transformação linear positiva, i.e.,

$$\rho_{a+bX, c+dY} = \rho_{X,Y} \quad \text{se } bd > 0 \quad (2.11)$$

Dem:

$$\rho_{a+bX, c+dY} = \frac{Cov[a + bX, c + dY]}{\sigma_{a+bX} \sigma_{c+dY}} = \frac{bd Cov[X, Y]}{|bd| \sigma_X \sigma_Y} = \frac{Cov[X, Y]}{\sigma_X \sigma_Y} = \rho_{X,Y}.$$

Exemplo 2.13 Estudo da variável média de n variáveis aleatórias independentes

Dadas n v.a. independentes e semelhantes, seja μ e σ^2 o valor médio e variância comuns.

A v.a. assim definida

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_i^n X_i,$$

chama-se **média** das n variáveis aleatórias.

Calculemos então o valor médio e a variância desta v.a., \bar{X}_n :

$$E[\bar{X}_n] = \frac{1}{n} E[X_1 + X_2 + \dots + X_n] = \frac{1}{n} n\mu = \mu$$

$$Var[\bar{X}_n] = \frac{1}{n^2} Var(X_1 + X_2 + \dots + X_n) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Função geradora de momentos

Até aqui estivemos a calcular duas características numéricas relevantes para o estudo de uma variável aleatória: o valor médio e a variância. Há situações em que os valores esperados a calcular, $E[X]$ e $E[X^2]$ são complexos. Um procedimento alternativo para o seu cálculo consiste em utilizar uma função chamada **função geradora de momentos** que, além de permitir obter aqueles valores esperados, tem a importante propriedade de caracterizar univocamente uma v.a., sempre que é possível aplicar-se. Mais adiante veremos outras propriedades extremamente importantes que aquela função possui.

Definição 2.16

Chama-se **função geradora de momentos (f.g.m.)** da v.a. X e representa-se por $M_X(t)$ a função assim definida

$$M_X(t) = E[e^{tX}] \quad (2.12)$$

quando o segundo membro da igualdade existe numa vizinhança de $t = 0$.

Tem-se $M_X(0) = 1$.

Sendo assim

$$M_X(t) = \begin{cases} \sum_i e^{tx_i} p_i & \text{se } X \text{ v.a. discreta} \\ \int_{-\infty}^{+\infty} e^{tx} f(x) dx & \text{se } X \text{ v.a. contínua.} \end{cases}$$

A função geradora de momentos de um par aleatório define-se como

$$M_{X,Y}(s, t) = E[e^{sX+tY}] \quad \text{e} \quad M_{X,Y}(0, 0) = 1. \quad (2.13)$$

O nome de função geradora de momentos justifica-se pelo teorema, que apresentamos sem demonstração:

Teorema 2.6 – (existência)

Se a função geradora de momentos está definida numa vizinhança de 0 tem-se

$$M_X^{(r)}(0) = E[X^r] \quad r = 1, 2, 3, \dots$$

onde $M_X^{(r)}(0)$ designa a r -ésima derivada de M_X calculada em $t = 0$.

Note-se que, em particular, o teorema nos diz que:

$$M_X'(0) = E[X] \quad \text{e} \quad M_X''(0) = E[X^2]$$

Exemplo 2.14

Consideremos a v.a. contínua com fdp dada por

$$f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

A f.g.m. é

$$M_X(t) = \int_0^{+\infty} e^{tx} e^{-x} dx = \int_0^{+\infty} e^{-x(1-t)} dx = \frac{1}{1-t} \quad \text{para } t < 1.$$

Exercício 2.7

Usando a f.g.m. calcule o valor médio e a variância da v.a. definida no exemplo anterior.

Propriedades da função geradora de momentos

1. $M_{a+bX}(t) = e^{at} M_X(bt)$.

Dem:

$$M_{a+bX}(t) = E[e^{t(a+bX)}] = E[e^{at} e^{btX}] = e^{at} M_X(bt).$$

2. Se X e Y são v.a. independentes

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

Dem:

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX} e^{tY}].$$

Tendo em conta o teorema 2.5, se X e Y são independentes também e^{tX} e e^{tY} são v.a. independentes. Pela propriedade 5. do valor médio

$$M_{X+Y}(t) = E[e^{tX}] E[e^{tY}] = M_X(t) M_Y(t).$$

Nota: A recíproca é mais fraca, diz-nos que se

$$M_{X+Y}(t) = M_X(t) M_Y(t)$$

então $E[XY] = E[X] E[Y]$, i.e., $Cov[X, Y] = 0$.

3. As variáveis aleatórias X e Y são independentes se e só se

$$M_{X,Y}(t_1, t_2) = M_{X,Y}(t_1, 0) M_{X,Y}(0, t_2)$$

A demonstração é deixada como exercício.

4. Teorema 2.7. - Teorema da unicidade

Se para duas v.a. X e Y se verifica $M_X(t) = M_Y(t)$ então X e Y são identicamente distribuídas. Reciprocamente, se existir a função geradora de momentos, ela é única.

Este teorema é apresentado sem demonstração.

Exercício 2.8

Usando propriedades da f.g.m. mostre que se X_1, \dots, X_n são n v.a. independentes e identicamente distribuídas com função geradora de momentos comum $M(t)$, então as funções geradoras de momentos de $S_n = \sum_i^n X_i$ e $\bar{X}_n = \sum_i^n X_i/n$ são respectivamente $M^n(t)$ e $M^n(t/n)$.

Principais distribuições discretas

Nesta secção iremos apresentar as distribuições discretas mais usadas como modelos de probabilidade discretos em muitas áreas de aplicação e principalmente nas áreas das Ciências da Vida.

A distribuição uniforme discreta

É a mais simples de todas as distribuições de probabilidade discretas – aquela em que a v.a. assume todos os seus valores com igual probabilidade.

Definição 2.17

Uma v.a. X diz-se ter **distribuição uniforme discreta** se toma os valores x_1, \dots, x_k com probabilidades $1/k, \dots, 1/k$, i.e.

$$P(X = x_i) = 1/k, \quad i = 1, \dots, k \quad (2.14)$$

Valor médio, variância e função geradora de momentos

$$E[X] = \mu = \frac{1}{k} \sum_{i=1}^k x_i; \quad Var[X] = \frac{1}{k} \sum_{i=1}^k (x_i - \mu)^2; \quad M_X(t) = \frac{1}{k} \sum_{i=1}^k e^{tx_i}.$$

O caso mais importante é o caso desta variável tomar os valores $1, 2, \dots, n$ (como por exemplo a v.a. que representa o número de face saída no lançamento de um dado perfeito). Neste caso tem-se²,

$$E[X] = \frac{1 + 2 + \dots + n}{n} = \frac{n+1}{2} \quad E[X^2] = \frac{(n+1)(2n+1)}{6}$$

portanto

$$Var[X] = E[X^2] - E^2[X] = \frac{n^2 - 1}{12}$$

$$M_X(t) = E[e^{tX}] = \frac{1}{n}(e^t + e^{2t} + \dots + e^{nt}) = \frac{e^t(1 - e^{nt})}{n(1 - e^t)}.$$

²Atenda-se a que a soma dos n primeiros números inteiros $1 + 2 + \dots + n = \frac{n(n+1)}{2}$ e que a soma dos quadrados dos n primeiros inteiros $1 + 4 + 9 + \dots + n^2 = \frac{n(n+1)(2n+1)}{6}$.

Distribuição binomial

Consideremos uma experiência na qual se observa a realização ou não de um dado acontecimento A . A realização de A costuma designar-se por sucesso sendo a realização do acontecimento complementar \bar{A} o insucesso. Estamos então a considerar experiências que têm apenas dois resultados possíveis.

Vejam alguns exemplos de tais experiências:

- o teste de uma dada droga num rato e o registo da reacção positiva ou negativa;
- a inspecção dos items numa linha de fabrico para observar se cada um é defeituoso ou não;
- o lançamento de uma moeda.

Suponhamos que fazemos repetições sucessivas de uma experiência nas condições anteriores. Cada repetição costuma chamar-se uma **prova**. Provas repetidas que verificam as condições seguintes:

- cada prova tem apenas um de dois resultados possíveis: sucesso ou insucesso.
- em cada prova a probabilidade de sucesso, p , permanece constante, sendo de $q = 1 - p$ a probabilidade de insucesso.
- as provas são independentes.

são chamadas **provas de Bernoulli**.

Considerando uma variável aleatória associada ao resultado de cada prova, i.e., tomando o valor 1 com probabilidade p se se verificar o sucesso e o valor 0 com probabilidade $1 - p = q$ se o resultado for insucesso, diz-se que a variável assim definida tem **distribuição de Bernoulli**.

Suponhamos que realizamos sucessivamente n provas de Bernoulli.

Definição 2.18

A v.a. X que conta o número de sucessos em n provas de Bernoulli chama-se **variável aleatória binomial**.

A distribuição de probabilidades da v.a. binomial chama-se **distribuição binomial** e representa-se por

$$X \sim \mathcal{B}(n, p)$$

para indicar que X tem distribuição binomial de parâmetros n (número de provas realizadas) e p (probabilidade constante de sucesso em cada prova).

A distribuição binomial caracteriza tiragens com reposição, quando há apenas dois resultados possíveis.

Exemplo 2.15

Consideremos a experiência que consiste no lançamento de uma moeda equilibrada. Cada lançamento (prova) tem dois resultados possíveis (coroa e face).

Consideremos sucesso – a saída de face e insucesso – a saída de coroa. Sendo a moeda equilibrada, tem-se $p = 1/2$ e $q = 1/2$.

Suponhamos que lançamos 5 vezes a moeda. Qual a probabilidade de que saia 3 vezes face?

Vejam os de quantos modos se pode observar três vezes a saída de face :

$$\begin{array}{ccccccccc} FFFCC & FFCFC & FFCCF & FCCFF & FCFCF \\ FCFFC & CFFFC & CCFFF & CFCFF & CFFCF \end{array}$$

O total de modos em que se observou 3 vezes face é $\binom{5}{3} = 10$.

Por outro lado e, dado que as provas são independentes e de probabilidade constante p , tem-se

$$P(FFFCC) = P(FFCFC) = \dots = P(FCFFC) = p^3 q^2 = \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2.$$

Sendo assim

$$P[X = 3] = P[(FFFCC) \cup (FFCFC) \cup \dots \cup (FCFFC)] = \binom{5}{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2$$

Podemos então caracterizar a v.a. $X \sim \mathcal{B}(n, p)$ do seguinte modo:

– toma os valores $x = 0, 1, 2, \dots, n$ com probabilidades assim definidas:

$$P[X = x] = \binom{n}{x} p^x (1 - p)^{n-x}. \quad (2.15)$$

Observe-se que X toma os valores $x = 0, 1, 2, \dots, n$, com probabilidades, respectivamente iguais a

$$(1 - p)^n, \quad np(1 - p)^{n-1}, \quad \binom{n}{2} p^2 (1 - p)^{n-2}, \quad \dots \quad p^n,$$

que são afinal os termos sucessivos do desenvolvimento do binómio $[p + (1 - p)]^n$; daí o nome de distribuição binomial.

Sendo assim tem-se

$$\sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1.$$

Na seguinte figura estão representados os gráficos de algumas probabilidades binomiais para $n = 5$ e vários valores de p .

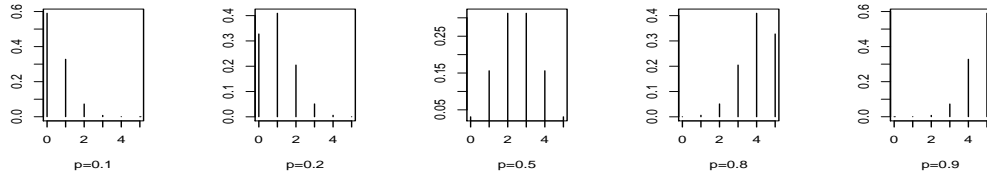


Figura 13: Gráficos da função massa de probabilidade de uma v.a. com distribuição $\mathcal{B}(5, p)$, para alguns valores de p .

O cálculo das diferentes probabilidades para vários valores de n e p é fastidioso e podemos consultar as tabelas com os resultados de muitos desses cálculos, sendo as mais vulgares construídas para vários valores de p e com n até 20 ou 25.

Algumas tabelas apresentam valores de p até 0.50. Se for necessário calcular probabilidades para valores de $p > 0.50$, basta ter em conta a relação

$$X \sim \mathcal{B}(n, p) \Rightarrow (n - X) \sim \mathcal{B}(n, 1 - p). \tag{2.16}$$

Exercício 2.9

A probabilidade de que um doente recupere de uma doença rara no sangue é de 0.4. Se soubermos que 15 pessoas contraíram a referida doença, qual é a probabilidade de que sobrevivam:

1. pelo menos 10;
2. entre 3 e 8;
3. exactamente 5.

Se X é a v.a. que designa o número de pessoas que sobrevivem, tem-se $X \sim \mathcal{B}(15; 0.4)$

1. $P[X \geq 10] = 1 - P[X \leq 9] = 1 - 0.966 = 0.034$;

2. $P[3 \leq X \leq 8] = 0.878;$

3. $P[X = 5] = 0.186.$

Valor médio, variância da distribuição binomial

Seja $X \sim \mathcal{B}(n, p).$

Consideremos a v.a. que designa o resultado da j -ésima prova, $I_j,$

$$I_j = \begin{cases} 1 & \text{com prob. } p, \text{ i.e. houve } \underline{\text{sucesso}} \\ 0 & \text{com prob. } q, \text{ i.e. houve } \underline{\text{insucesso}}. \end{cases}$$

I_j é então uma **variável de Bernoulli**, também designada por **variável indicatriz**.

Sendo assim, numa experiência binomial com n provas, o número de sucessos pode ser escrito como a soma de n variáveis indicatrizes independentes:

$$X = I_1 + I_2 + \dots + I_n.$$

Calculemos o valor médio e a variância da v.a. $I_j:$

$$E[I_j] = 1 \cdot p + 0 \cdot q = p; \quad \text{Var}[I_j] = E[I_j^2] - [E[I_j]]^2 = p - p^2 = p(1 - p) = pq$$

donde

$$E[X] = E[I_1 + I_2 + \dots + I_n] = E[I_1] + E[I_2] + \dots + E[I_n] = np$$

$$\text{Var}[X] = \text{Var}[I_1 + \dots + I_n] = \text{Var}[I_1] + \text{Var}[I_2] + \dots + \text{Var}[I_n] = npq$$

atendendo a que I_j são variáveis independentes.

Função geradora de momentos

Começemos por calcular $M_{I_j}(t) = E[e^{tI_j}] = e^t p + e^0 q = e^t p + q.$

Como as v.a. I_j são independentes, pela propriedade 2. da função geradora de momentos

$$M_X(t) = M_{I_1+I_2+\dots+I_n}(t) = (p e^t + q)^n. \tag{2.17}$$

Exercício 2.10

Usando a função geradora de momentos obter o valor médio e a variância da v.a. $X \sim \mathcal{B}(n, p).$

Distribuição multinomial

Esta distribuição generaliza a binomial. Dada uma experiência em que cada prova pode ter mais de dois resultados possíveis, verificando todas as outras condições referidas na distribuição binomial, diz-se que estamos perante uma **experiência multinomial**.

Como exemplo de experiências multinomiais temos o caso de tiragens com reposição de cartas de um baralho.

Suponhamos então que cada prova tem k resultados possíveis:

E_1, E_2, \dots, E_k com probabilidades
 p_1, p_2, \dots, p_k $p_i \geq 0$ e $\sum_{i=1}^k p_i = 1$.

Pretendemos determinar a probabilidade de em n provas independentes observar x_1 vezes o acontecimento E_1
 x_2 vezes o acontecimento E_2
...
 x_k vezes o acontecimento E_k , com $x_1 + x_2 + \dots + x_k = n$.

Sejam X_1, X_2, \dots, X_k as variáveis aleatórias que designam o número de vezes que sai cada um dos acontecimentos nas n provas. Então, atendendo a que $\frac{n!}{x_1! x_2! \dots x_k!}$ são todos os modos possíveis de que seja verificada tal ocorrência tem-se

$$P[X_1 = x_1, X_2 = x_2, \dots, X_k = x_k] = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}.$$

Exemplo 2.16

De acordo com a teoria genética, um certo cruzamento de porcos da índia resultará em descendência vermelha, preta e branca na proporção de 8:4:4.

Determine a probabilidade de que, entre 8 descendentes, 5 sejam vermelhos, 2 pretos e 1 branco.

Tem-se $p_1 = 8/16 = 1/2$ $p_2 = 1/4$ $p_3 = 1/4$ donde

$$P[X_1 = 5, X_2 = 2, X_3 = 1] = \frac{8!}{5! 2! 1!} (1/2)^5 (1/4)^2 (1/4).$$

A distribuição binomial negativa

A distribuição binomial estudada umas páginas atrás, é o modelo probabilístico adequado para descrever os resultados associados a uma sucessão de provas independentes em que, em cada uma, se verifica ou não a realização de um dado acontecimento A . A distribuição binomial supõe a realização de n provas independentes, sendo aleatório o número de "sucessos" observados nessas n provas.

Considere-se agora que se fixa o número de “sucessos”, k e se pretende contar o **número de provas independentes necessárias até obter aqueles k “sucessos”**. Neste caso o **número de provas é aleatório**.

A variável, X assim definida, diz-se ter **distribuição binomial negativa** e é costume representar-se por $X \sim \mathcal{BN}(k, p)$, onde p é a probabilidade constante de “sucesso” de prova para prova e k é o número de “sucessos”.

A função massa de probabilidade é assim definida

$$P[X = x] = P \left[\begin{array}{l} \text{em } (x - 1) \text{ provas independentes haver } (k - 1) \text{ sucessos} \\ \text{e na } x.\text{ésima prova realizar-se o } r.\text{ésimo sucesso} \end{array} \right]$$

$$= \binom{x-1}{k-1} p^{k-1} q^{x-1-(k-1)} p = \binom{x-1}{k-1} p^k q^{x-k}$$

$$\text{com } x = k, k + 1, \dots \quad 0 < p < 1, \quad q = 1 - p.$$

Valor médio, variância e função geradora de momentos de $X \sim \mathcal{BN}(k, p)$

$$E[X] = \frac{k}{p} \quad \text{Var}[X] = \frac{kq}{p^2} \quad M_X(t) = \left(\frac{p e^t}{1 - qe^t} \right)^k \quad \text{com } qe^t < 1$$

Um caso particular desta distribuição ocorre fazendo $k = 1$, obtendo-se a chamada **distribuição geométrica**.

Se $X \sim \mathcal{BN}(1, p)$ é costume representar-se por $X \sim \mathcal{G}(p)$

X designa o **número de provas necessárias até que ocorra o primeiro “sucesso”**.

A função massa de probabilidade é assim definida

$$P[X = x] = pq^{x-1} \quad x = 1, 2, \dots \quad 0 < p < 1 \quad q = 1 - p$$

Valor médio, variância e função geradora de momentos de $X \sim \mathcal{G}(p)$

$$M_X(t) = \frac{p e^t}{1 - qe^t} \quad qe^t < 1; \quad E[X] = 1/p; \quad \text{Var}[X] = q/p^2$$

Propriedade da falta de memória, da distribuição geométrica:

Teorema 2.8

Se $X \sim \mathcal{G}(p)$ então sendo m e n inteiros positivos

$$P[X > m + n | X > m] = P[X > n]$$

Nota: Interpretando a distribuição geométrica como o tempo de espera por um “sucesso”: se tiverem decorrido mais de m provas sem que se tenha verificado um “sucesso”, a probabilidade de se ter de esperar mais de n provas para se observar um “sucesso” é a mesma que esperar mais de n provas caso se estivesse no início da experiência

Observação: O recíproco deste teorema também é verdadeiro.

Distribuição hipergeométrica

Vimos que a distribuição binomial caracterizava tiragens com reposição. Porém, experiências que consistem em tiragens sem reposição, violam as condições exigidas em provas de Bernoulli.

Consideremos o seguinte

Exemplo 2.17

De um baralho de 52 cartas, qual é a probabilidade que um jogador tem de seleccionar 3 vermelhas e 2 pretas (note-se que ao retirar cartas para jogar não pode haver reposição).

Como temos 26 cartas vermelhas e 26 cartas pretas, há $\binom{26}{3}$ modos de seleccionar 3 cartas vermelhas e, para cada um, há $\binom{26}{2}$ modos de escolher 2 cartas pretas.

Sendo assim, o número total de modos de seleccionar 3 cartas vermelhas e 2 pretas é $\binom{26}{3}\binom{26}{2}$ enquanto o número total de modos de seleccionar 5 cartas quaisquer do baralho é $\binom{52}{5}$.

Portanto a probabilidade de seleccionar 5 cartas, sem reposição, sendo 3 vermelhas e 2 pretas é:

$$\frac{\binom{26}{3}\binom{26}{2}}{\binom{52}{5}}$$

Regra geral estamos interessados em calcular a probabilidade de seleccionar x elementos de entre K items que consideramos “sucessos” e $n - x$ elementos de entre $N - K$ items rotulados de “insucessos”, quando extraímos uma amostra aleatória de n elementos dos N items.

Temos uma **experiência hipergeométrica** quando temos a seguinte situação:

- de N elementos seleccionamos n elementos ao acaso;

- K desses N elementos são classificados de sucessos e $N - K$ são classificados de insucessos.

Seja X a v.a. que conta o número de sucessos numa prova hipergeométrica, diz-se então que X é uma v.a. hipergeométrica de parâmetros N , n e K e costuma representar-se por

$$X \sim \mathcal{H}(N, n, K)$$

Para esta v.a. tem-se a distribuição de probabilidades

$$P[X = x] = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}} \quad (2.18)$$

$$\max(0, n - N + K) \leq x \leq \min(n, K)$$

Valor médio e variância da distribuição hipergeométrica

No que se segue iremos supor $0 \leq x \leq n$, i.e., $n \leq K$ e $n \leq N - K$.
Atendendo a que

$$\sum_{x=0}^n \binom{K}{x} \binom{N-K}{n-x} = \binom{K+N-K}{x+n-x} = \binom{N}{n}$$

é fácil verificar que

$$\sum_{x=0}^n \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}} = 1, \quad \text{tratando-se por isso de uma distribuição de probabilidades.}$$

$$E[X] = \sum_{x=0}^n x \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}} =$$

$$= 0 + 1 \frac{\binom{K}{1} \binom{N-K}{n-1}}{\binom{N}{n}} + 2 \frac{\binom{K}{2} \binom{N-K}{n-2}}{\binom{N}{n}} + \dots + n \frac{\binom{K}{n} \binom{N-K}{0}}{\binom{N}{n}} = \sum_{x=1}^n x \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}} = *$$

Atendendo a que $x \binom{K}{x} = K \binom{K-1}{x-1}$ tem-se

$$* = K \sum_{x=1}^n \frac{\binom{K-1}{x-1} \binom{N-K}{n-x}}{\binom{N}{n}} = K \sum_{y=0}^{n-1} \frac{\binom{K-1}{y} \binom{N-K}{n-y-1}}{\binom{N}{n}} \quad \text{com } y = x - 1.$$

Tendo em conta que

$$\binom{N-K}{n-y-1} = \binom{(N-1)-(K-1)}{n-y-1} \quad \text{e que} \quad \binom{N}{n} = \frac{N}{n} \binom{N-1}{n-1}$$

temos finalmente

$$E[X] = \frac{nK}{N} \sum_{y=0}^{n-1} \frac{\binom{K-1}{y} \binom{(N-1)-(K-1)}{n-y-1}}{\binom{N-1}{n-1}}.$$

Ora este somatório representa a soma de todas as probabilidades duma v.a. $\mathcal{H}(N-1, n-1, K-1)$, portanto igual a 1. Donde

$$E[X] = n \frac{K}{N}.$$

Exercício 2.11

Considerando procedimentos análogos aos acabados de apresentar mostrar que

$$Var[X] = n \frac{K}{N} \left(1 - \frac{K}{N}\right) \frac{N-n}{N-1}.$$

Observação: Quando N é grande comparado com n , a probabilidade de sucesso em cada tiragem sem reposição varia muito pouco de prova para prova, por conseguinte esbate-se a diferença entre tiragens sem e com reposição. Sendo assim, a distribuição binomial aproxima a distribuição hipergeométrica com $p = K/N$.

Este resultado resulta do seguinte teorema que se enuncia sem demonstração:

Teorema 2.9.

Com n e K/N fixos

$$\lim_{N \rightarrow \infty} \left[\frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}} \right] = \binom{n}{x} (K/N)^x (1 - K/N)^{n-x}.$$

Conclusão: Se n bastante menor que N tem-se

$$\mathcal{H}(N, n, K) \approx \mathcal{B}(n, K/N).$$

Como regra prática, pode considerar-se boa a aproximação para $n < N/10$.

Observe-se que as fórmulas do valor médio e variância de cada uma das distribuições estão relacionadas, em face da aproximação estabelecida pelo teorema 2.9:

	<u>Binomial</u>	<u>Hipergeométrica</u>
valor médio	np	nK/N
variância	npq	$n\frac{K}{N}(1 - \frac{K}{N})\frac{N-n}{N-1}$
		o factor $\frac{N-n}{N-1}$ chama-se factor de correcção que se torna desprezável quando $n \ll N$.

A distribuição hipergeométrica generalizada³

A distribuição hipergeométrica pode ser estendida ao caso de termos N elementos repartidos por K células A_1, A_2, \dots, A_K , cada uma com a_1, a_2, \dots, a_K elementos.

Retirada uma amostra sem reposição de dimensão n de entre os N elementos, pretendemos determinar a probabilidade de que tenhamos

x_1 elementos de A_1

x_2 elementos de A_2

...

x_K elementos de A_K com $x_1 + x_2 + \dots + x_K = n$

O valor dessa probabilidade é

$$\frac{\binom{a_1}{x_1} \binom{a_2}{x_2} \dots \binom{a_K}{x_K}}{\binom{N}{n}}$$

Exemplo 2.18

Uma urna contém 8 bolas vermelhas, 6 bolas brancas e 10 bolas pretas. Qual a probabilidade de ao retirar 6 bolas sem reposição, obter 3 bolas vermelhas, 1 branca e 2 pretas?

Resolução:

$$\frac{\binom{8}{3} \binom{6}{1} \binom{10}{2}}{\binom{24}{6}}.$$

Distribuição de Poisson

As experiências que conduzem a valores numéricos de uma v.a. X que conta o número de “sucessos” que ocorrem num dado intervalo de tempo ou num domínio específico, satisfazendo certas condições, são designadas por **experiências de Poisson**.

³Esta distribuição não será referida este ano lectivo.

A distribuição de Poisson constitui um modelo matemático de fenómenos aleatórios, se são verificadas as seguintes condições:

- o número de sucessos que ocorre num dado intervalo de tempo ou domínio é independente do número que ocorre em qualquer outro intervalo ou domínio disjunto dos anteriores;
- a probabilidade que o acontecimento se realize uma vez em qualquer intervalo muito curto (ou região muito pequena) é proporcional à amplitude do intervalo ou tamanho da região e não depende do número de sucessos que ocorrem fora desse intervalo ou região;
- a probabilidade de que o acontecimento se realize mais do que uma vez num intervalo de amplitude muito pequena é desprezável.

Muitos fluxos de acontecimentos observam, pelo menos com grande aproximação, as hipóteses enunciadas. São exemplos:

- chamadas telefónicas recebidas numa dada central telefónica num certo intervalo de tempo;
- chegada de clientes à bilheteira de uma estação de caminhos de ferro;
- chegada de sinistrados a um banco de um hospital;
- número de dias que uma dada escola fecha durante o inverno;
- número de erros de tipografia por página;
- emissão de partículas α , que são detectadas num contador.

Definição 2.19

A v.a X que conta o número de sucessos numa experiência de Poisson diz-se ter **distribuição de Poisson** e prova-se que depende apenas de um parâmetro λ , que representa o número médio de sucessos que ocorrem no intervalo de tempo (ou na região especificada). Costuma representar-se por

$$X \sim \mathcal{P}(\lambda)$$

Não apresentaremos aqui a dedução da distribuição de probabilidades da v.a. X , por constituir matéria fora do conhecimento dos alunos. Tem-se

$$P[X = x] = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots \quad (2.19)$$

onde λ designa o número médio de sucessos que ocorrem no intervalo de tempo considerado ou na região especificada.

Tem-se

$$P[X = x] \geq 0 \quad \forall x = 0, 1, 2, \dots$$

e a teoria das séries de funções permite-nos garantir que

$$\sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1.$$

Trata-se portanto de uma distribuição de probabilidades discreta.

Existem também tabelas para o cálculo destas probabilidades, regra geral para um domínio de valores de λ entre 0.1 e 20.

Exemplo 2.19

O número médio de petroleiros que chegam por dia a um dado porto é 10. As instalações do porto conseguem abrigar no máximo 15. Qual a probabilidade que num dado dia se tenha de recusar abrigo a petroleiros?

Resolução: Tem-se então $X \sim \mathcal{P}(10)$ e pretende-se determinar

$$P[X > 15] = 1 - P[X \leq 15] = 1 - \sum_{x=0}^{15} \frac{e^{-10} 10^x}{x!} = 1 - 0.9513 = 0.0487.$$

por leitura directa nas tabelas da Poisson cumulativa.

Valor médio, variância e função geradora de momentos.

Comecemos por calcular a função geradora de momentos, $M_X(t)$

$$M_X(t) = E[e^{tX}] = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t-1)}. \quad (2.20)$$

$$E[X] = \left. \frac{d M_X(t)}{dt} \right|_{t=0} = \left. \frac{d e^{\lambda(e^t-1)}}{dt} \right|_{t=0} = \left. \lambda e^t e^{\lambda(e^t-1)} \right|_{t=0} = \lambda.$$

$$E[X^2] = \left. \frac{d^2 M_X(t)}{dt^2} \right|_{t=0} = \left. \lambda e^t e^{\lambda(e^t-1)} + \lambda e^t \lambda e^t e^{\lambda(e^t-1)} \right|_{t=0} = \lambda + \lambda^2.$$

Portanto

$$\sigma_X^2 = E[X^2] - \lambda^2 = \lambda.$$

Exercício 2.12

Deduzir, usando directamente a definição, o valor médio e a variância de $X \sim \mathcal{P}(\lambda)$.

Teorema 2.10

Se as v.a. X_i $i = 1, \dots, k$ são independentes e $X_i \sim \mathcal{P}(\lambda_i)$ então

$$\sum_{i=1}^k X_i \sim \mathcal{P}\left(\sum_{i=1}^k \lambda_i\right).$$

Dem: é imediata recorrendo à f.g.m. e às suas propriedades:

$$M_{\sum_{i=1}^k X_i}(t) = M_{X_1}(t) M_{X_2}(t) \dots M_{X_k}(t) = e^{\lambda_1(e^t-1)} e^{\lambda_2(e^t-1)} \dots e^{\lambda_k(e^t-1)} = \exp\left(\sum_{i=1}^k \lambda_i(e^t-1)\right)$$

que é a f.g.m. de uma v.a. com distribuição de Poisson de parâmetro $\sum_{i=1}^k \lambda_i$.

A distribuição de Poisson pode ainda ser obtida como limite da distribuição binomial quando $n \rightarrow \infty$ e $p \rightarrow 0$.

Teorema 2.11

A distribuição binomial converge para a distribuição de Poisson quando $n \rightarrow \infty$, $p \rightarrow 0$, mantendo-se constante o produto np , ou seja, tem-se nas condições referidas

$$X \sim \mathcal{B}(n, p) \approx X \sim \mathcal{P}(np).$$

Dem: Fazendo $p = \lambda/n$ tem-se

$$\begin{aligned} \binom{n}{x} p^x (1-p)^{n-x} &= \frac{n!}{x!(n-x)!} (\lambda/n)^x (1-\lambda/n)^{n-x} = \\ &= \frac{n(n-1)\dots(n-x+1)}{n^x} \frac{\lambda^x}{x!} (1-\lambda/n)^{-x} (1-\lambda/n)^n, \end{aligned}$$

como o primeiro e terceiro factores tendem para 1, o segundo é constante em n e o quarto tem como limite $e^{-\lambda}$, vem

$$\lim_{n \rightarrow \infty} \binom{n}{x} p^x (1-p)^{n-x} = \frac{e^{-\lambda} \lambda^x}{x!}.$$

Em geral, a distribuição de Poisson fornece uma boa aproximação da distribuição binomial quando $n \geq 20$ e $p \leq 0.05$; quando $n \geq 100$ e $np \leq 10$ a aproximação é muito boa.

Nota: A distribuição de Poisson também se chama **a lei dos fenômenos raros** por se aplicar a problemas em que a probabilidade p de ocorrência do acontecimento é muito pequena em comparação com o número de observações n .

Principais distribuições contínuas

Nesta secção iremos apresentar três distribuições contínuas, que habitualmente são tratadas em cursos introdutórios, como este que estamos a realizar.

No final deste capítulo deixamos o estudo de mais alguns modelos, que são instrumentos fundamentais na modelação e/ou estimação de características nas mais variadas áreas de aplicação.

A distribuição uniforme contínua

Uma v.a. contínua diz-se que tem **distribuição uniforme** ou **rectangular** no intervalo (a, b) e representa-se simbolicamente por $X \sim \mathcal{U}(a, b)$ se a função densidade de probabilidade (f.d.p.) é da forma:

$$f(x) = \begin{cases} 1/(b-a) & a < x < b \\ 0 & x \leq a \text{ ou } x \geq b. \end{cases}$$

A correspondente função de distribuição é:

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

Valor médio, variância e função geradora de momentos.

$$E[X] = \int_{-\infty}^{+\infty} xf(x)dx = \int_a^b x \cdot 1/(b-a)dx = (a+b)/2;$$

$$Var[X] = E[X^2] - E^2[X] = \int_a^b x^2 \cdot 1/(b-a)dx - (a+b)^2/4 = (b-a)^2/12;$$

$$M_X(t) = E[e^{tx}] = (e^{tb} - e^{ta})/t(b-a) \quad t \neq 0.$$

O caso particular da distribuição uniforme $\mathcal{U}(0, 1)$ é o que apresenta mais interesse, devido ao seguinte teorema:

Teorema 2.12 – Transformação uniformizante

Seja X uma v.a. contínua, com função de distribuição $F(x)$. Então a v.a. $Y = F(X) \sim \mathcal{U}(0, 1)$.

Dem: A nova v.a. Y , dado o modo como foi definida e pelas propriedades da função de distribuição, toma valores no intervalo $(0, 1)$.

Calculemos agora a sua função de distribuição:

$$G(y) = P[Y \leq y] = P[F(X) \leq y] = P[X \leq F^{-1}(y)],$$

visto $F(x)$ ser uma função monótona não decrescente, logo invertível. Tem-se então

$$G(y) = \begin{cases} 0 & y \leq 0 \\ P[X \leq F^{-1}(y)] = F(F^{-1}(y)) = y & 0 < y < 1 \\ 1 & y \geq 1 \end{cases},$$

que é a função de distribuição de uma v.a. $\mathcal{U}(0, 1)$.

A Distribuição Normal ou de Gauss

A **distribuição normal** surgiu no século XVIII ligada ao estudo dos erros de medições repetidas de uma mesma quantidade. As suas propriedades matemáticas foram estudadas por De Moivre, Laplace e Gauss, em honra de quem também se costuma chamar a esta distribuição, **distribuição de Gauss**.

É uma das mais importantes distribuições contínuas, sendo variadas as razões que para isso contribuem:

- muitas variáveis biométricas têm uma distribuição muito próxima da normal;
- por vezes uma variável que não é normal pode ser transformada de um modo simples numa outra com distribuição normal;
- a parte central de muitos modelos não normais é por vezes razoavelmente bem aproximada por uma distribuição normal.

Definição 2.20

Uma v.a. contínua X diz-se ter uma **distribuição normal** com parâmetros μ e σ e representa-se por $\mathbf{X} \sim \mathcal{N}(\mu, \sigma)$ se a sua f.d.p. é da forma:

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right] \quad (2.21)$$

$$-\infty < x < +\infty, \quad -\infty < \mu < +\infty, \quad 0 < \sigma < +\infty$$

Vejamos em primeiro lugar que de facto $f(x)$ é uma função densidade:

- É imediato que $f(x) > 0$;

- Verifiquemos que:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Efectivamente tem-se

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi} \sigma} \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right] dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-y^2/2} dy$$

depois de efectuada a mudança de variável $y = (x - \mu)/\sigma$.

Como pretendemos provar que

$$A = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-y^2/2} dy = 1$$

consideremos

$$A^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-y^2/2} dy \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-u^2/2} du = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(y^2+u^2)/2} du dy.$$

Efectuando a mudança para coordenadas polares $y = \rho \operatorname{sen} \theta$ e $u = \rho \operatorname{cos} \theta$, vem

$$A^2 = \frac{1}{2\pi} \int_0^{2\pi} d\theta \int_0^{+\infty} \rho e^{-\rho^2/2} d\rho = 1.$$

Temos então verificado que efectivamente $f(x)$ definida em (2.21) é uma função densidade.

Propriedades da distribuição normal

1. A curva normal é simétrica relativamente a $x = \mu$.

Efectivamente é imediato verificar que $f(\mu + a) = f(\mu - a)$.

2. É uma curva unimodal, sendo a moda (ponto do eixo das abcissas em que ocorre o máximo) $x = \mu$.

Basta fazer o estudo da derivada

$$f'(x) = -\frac{1}{\sqrt{2\pi} \sigma^2} \left(\frac{x - \mu}{\sigma} \right) \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right].$$

É fácil verificar que $x = \mu$ é um maximizante de $f(x)$, sendo o valor máximo da função, $f(\mu) = 1/\sqrt{2\pi}\sigma$.

3. A curva normal tem pontos de inflexão em $x = \mu + \sigma$ e $x = \mu - \sigma$.

Trata-se de um exercício simples de estudo da segunda derivada, $f''(x)$.

Na figura seguinte pode ver-se alguns aspectos que a função densidade normal pode apresentar, dependendo apenas dos valores de μ e σ .

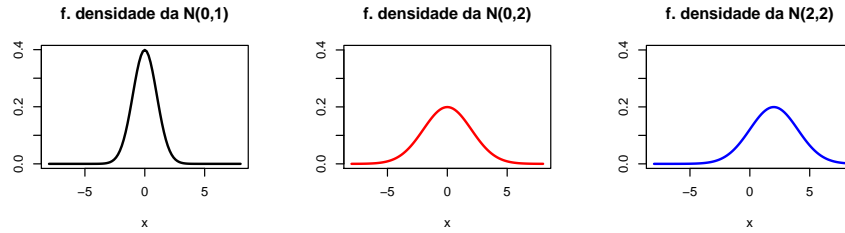


Figura 14: Gráficos da função densidade de $X \sim \mathcal{N}(0, 1)$, $X \sim \mathcal{N}(0, 2)$ e $X \sim \mathcal{N}(2, 1)$ (da esquerda para a direita)

Valor médio, variância e função geradora de momentos.

A **função geradora de momentos** é, como sabemos, assim definida

$$\begin{aligned}
 M_X(t) &= E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \\
 &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{-2t\sigma^2x + x^2 - 2x\mu + \mu^2}{2\sigma^2}} dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2 - 2(t\sigma^2 + \mu)x + \mu^2}{2\sigma^2}} dx = \\
 &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2 - 2(t\sigma^2 + \mu)x + (\mu^2 + 2t\sigma^2\mu + t^2\sigma^4) - 2t\sigma^2\mu - t^2\sigma^4}{2\sigma^2}} dx = \\
 &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left[\left(\frac{x - (t\sigma^2 + \mu)}{\sigma}\right)^2 - 2t\mu - t^2\sigma^2\right]} dx
 \end{aligned}$$

Fazendo a mudança de variável $\frac{x - (t\sigma^2 + \mu)}{\sigma} = y$, temos

$$M_X(t) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(y^2 - 2t\mu - t^2\sigma^2)} \sigma dy = e^{t\mu + \frac{t^2\sigma^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = e^{t\mu + \frac{t^2\sigma^2}{2}}$$

logo

$$M_X(t) = e^{t\mu + \frac{t^2\sigma^2}{2}} \quad (2.22)$$

Façamos agora o cálculo do valor médio e da variância:

$$\frac{dM_X(t)}{dt} = \left[\mu + \frac{2\sigma^2 t}{2} \right] e^{t\mu + \frac{t^2\sigma^2}{2}} \Big|_{t=0} = \mu$$

donde

$$E[X] = \mu \quad (2.23)$$

$$\frac{d^2 M_X(t)}{dt^2} = \sigma^2 e^{t\mu + \frac{t^2\sigma^2}{2}} + \mu[\mu + \sigma^2 t] e^{t\mu + \frac{t^2\sigma^2}{2}} \Big|_{t=0} = \sigma^2 + \mu^2$$

portanto

$$\text{Var}[X] = E[X^2] - [E[X]]^2 = \sigma^2 + \mu^2 - \mu^2, \quad \text{donde}$$

$$\text{Var}[X] = \sigma^2 \quad (2.24)$$

Exercício 2.13

Calcular o valor médio e a variância da v.a. normal, usando directamente a definição.

O cálculo das áreas sob a curva normal é um problema difícil de resolver por envolver integrais da função definida em (20). Há porém valores tabelados da função de distribuição da v.a. com $\mu = 0$ e $\sigma = 1$.

À variável aleatória com distribuição $\mathcal{N}(0, 1)$ chama-se **normal reduzida** e passaremos a representá-la por Z .

A função densidade da normal reduzida será representada por $\varphi(z)$ e é portanto assim definida

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad (2.25)$$

É imediato verificar que para a v.a. $Z \sim \mathcal{N}(0, 1)$ se tem

$$E[Z] = 0 \quad \text{Var}(Z) = 1 \quad M_Z(t) = e^{t^2/2}.$$

A leitura das tabelas da normal reduzida é fácil – em cada ponto a tabela dá-nos o valor da função de distribuição da normal reduzida (área sob a curva normal reduzida até ao ponto especificado).

Designemos por $\Phi(z) = P[Z \leq z]$, i.e., $\Phi(\cdot)$ é a função de distribuição da normal reduzida.

Observe-se que em consequência da simetria, se tem a seguinte relação:

$$\Phi(-z) = 1 - \Phi(z).$$

Consideremos o seguinte teorema que nos permite transformar uma v.a. normal qualquer numa v.a. normal reduzida.

Teorema 2.13

Seja $X \sim \mathcal{N}(\mu, \sigma)$ então a v.a. $Y = a + bX \sim \mathcal{N}(a + b\mu, |b|\sigma)$.

Dem: Recorrendo à f.g.m. da v.a. $X \sim \mathcal{N}(\mu, \sigma)$, que como vimos em (2.22) é dada por

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

e usando a propriedade 1. da função geradora de momentos,

$$M_{a+bX}(t) = e^{at} M_X(bt) = e^{at} e^{\mu bt + \frac{\sigma^2 b^2 t^2}{2}} = e^{(a + b\mu)t + \frac{(b\sigma)^2 t^2}{2}},$$

que é a f.g.m. de uma variável aleatória com distribuição normal de valor médio $(a + b\mu)$ e desvio padrão $|b|\sigma$.

Corolário

Seja $X \sim \mathcal{N}(\mu, \sigma)$ então a v.a. $\frac{X - \mu}{\sigma}$ tem distribuição normal reduzida, i.e., $Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$.

Dem: Para fazer a demonstração basta considerar no teorema acima $a = -\mu/\sigma$ e $b = 1/\sigma$.

Exercício 2.14

Dada $X \sim \mathcal{N}(1, 4)$, calcule:

1. $P(X < 6)$;
2. $P(-3 < X < 10)$;
3. $P(X > 0)$.

Resolução:

1. $P(X < 6) = P\left(\frac{X-1}{2} < \frac{6-1}{2}\right) = P(Z < 1.25) = \Phi(1.25) = 0.8944$;
2. $P(-3 < X < 10) = \Phi(2.25) - \Phi(-1) = 0.8291$;
3. $P(X > 0) = 1 - P(X \leq 0) = 1 - \Phi(-0.25) = \Phi(0.25) = 0.5987$.

Consideremos agora alguns teoremas de grande importância no estudo da distribuição normal.

Teorema 2.14 – Teorema da estabilidade da soma de normais.

Sejam X_1, \dots, X_n , v.a. normais independentes, tais que

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1)$$

$$X_2 \sim \mathcal{N}(\mu_2, \sigma_2)$$

...

$$X_n \sim \mathcal{N}(\mu_n, \sigma_n).$$

Então a v.a $X = X_1 + X_2 + \dots + X_n$ tem distribuição normal de parâmetros (μ, σ) ,

com

$$\mu = \mu_1 + \mu_2 + \dots + \mu_n \quad \text{e}$$

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}$$

Dem: A demonstração é muito fácil recorrendo de novo à f.g.m. Como sabemos

$$M_{X_i}(t) = \exp\left[\mu_i t + \frac{\sigma_i^2 t^2}{2}\right] \quad i = 1, \dots, n$$

então pela propriedade 2. da f.g.m. vem

$$M_X(t) = \prod_{i=1}^n M_{X_i}(t) = \exp\left[\sum \mu_i t + \sum \frac{\sigma_i^2 t^2}{2}\right].$$

Trata-se portanto da f.g.m. duma v.a. normal tendo como parâmetros $\mu = \sum \mu_i$ e $\sigma = \sqrt{\sum \sigma_i^2}$.

Exercício 2.15

Considerar uma generalização do teorema anterior provando que a distribuição de $X = a_1 X_1 + a_2 X_2 + \dots + a_n X_n$ é normal de valor médio e desvio padrão respectivamente

$$\mu = a_1 \mu_1 + a_2 \mu_2 + \dots + a_n \mu_n \quad \text{e}$$

$$\sigma = \sqrt{a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_n^2 \sigma_n^2}.$$

Corolário

Sejam X_i n v.a. normais independentes e semelhantes, i.e., tendo todas o mesmo valor médio μ e a mesma variância σ^2 .

A variáveis aleatórias soma e média, definidas respectivamente como

$$S = \sum_{i=1}^n X_i \quad \text{e} \quad \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

têm distribuição normal assim definida

$$S \sim \mathcal{N}(n\mu, \sigma\sqrt{n}) \quad \text{e} \quad \bar{X}_n \sim \mathcal{N}(\mu, \sigma/\sqrt{n}).$$

Deixamos como exercício a demonstração deste corolário.

No teorema anterior provámos que a soma de normais independentes é ainda uma normal. Porém um dos teoremas mais importantes de toda a teoria das probabilidades, o **teorema limite central**, dá-nos a distribuição aproximada da soma de n variáveis aleatórias independentes e identicamente distribuídas, desde que os momentos verifiquem certas condições. Vejamos então o enunciado do teorema, sem fazermos a sua demonstração.

Teorema 2.15 - Teorema Limite Central

Seja $\{X_n\}$ uma sucessão de n variáveis aleatórias independentes e identicamente distribuídas, com valor médio μ e variância σ^2 (finita). A v.a. $S_n = \sum_{i=1}^n X_i$, depois de estandardizada, i.e.

$$\frac{S_n - n\mu}{\sigma\sqrt{n}},$$

tem, assintoticamente, distribuição normal reduzida, i.e., para valores de n bastante ‘grandes’ tem-se

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

donde também se tem

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

Um teorema cuja demonstração se torna agora fácil é o teorema que De Moivre enunciou para variáveis aleatórias com distribuição Binomial.

Teorema 2.16 - Teorema de De Moivre

Seja X uma v.a. com distribuição binomial com valor médio $\mu = np$ e variância $\sigma^2 = npq$. Então quando $n \rightarrow \infty$,

$$\frac{X - np}{\sqrt{npq}}$$

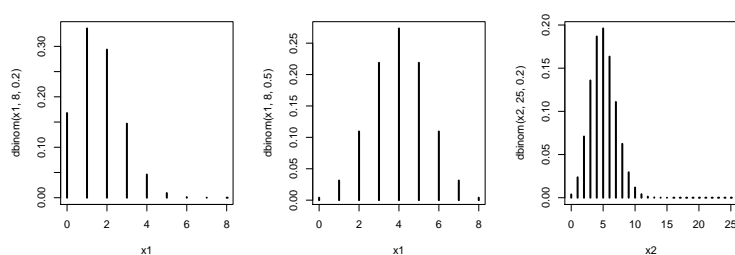
tem aproximadamente **distribuição normal reduzida**.

Dem: Como sabemos, a v.a. $X \sim B(n, p)$, pode escrever-se como a soma de n variáveis indicatrizes independentes, i.e., $X = I_1 + I_2 + \dots + I_n$, para as quais se tem $E[I_j] = p$ e $Var(I_j) = pq$; $j = 1, \dots, n$. É portanto imediato o resultado do teorema.

Tínhamos já visto que quando o parâmetro n da distribuição binomial era grande (não havia valores das probabilidades nas tabelas usuais), era necessário calcular as probabilidades binomiais por processos aproximados. Vimos então que a distribuição de Poisson podia ser usada como uma aproximação da distribuição binomial quando n grande e p próximo de zero ou um.

Porém ficávamos sem resposta ao caso de p tomar valores próximos de $1/2$. É para valores de p nestas condições que o teorema acabado de enunciar, oferece muito boa aproximação.

A figura seguinte ilustra que, mesmo para valores de n relativamente pequenos, a aproximação é bastante boa desde que $p \simeq 1/2$.



Gráficos da função massa de probabilidade de uma v.a. com distribuição $B(8, 0.2)$, $B(8, 0.5)$ e $B(25, 0.2)$, da esquerda para a direita, respectivamente.

Como regra prática podemos considerar boa a aproximação da distribuição binomial pela distribuição normal, quando $np > 5$ e $nq > 5$.

Um outro teorema, aplicação do teorema limite central é o que nos permite obter a lei limite de uma distribuição de Poisson. Como vimos já, a soma de v. a. de Poisson independentes é uma variável de Poisson. Sendo assim, em particular, toda a v.a. de Poisson de parâmetro m (inteiro) pode ser considerada uma soma de m v.a. de Poisson de valor médio 1. Daí resulta, por aplicação do teorema limite central, que quando $m \rightarrow \infty$ a distribuição de Poisson tende para a distribuição normal.

Teorema 2.17

Seja X uma v.a. com distribuição de Poisson de valor médio λ . Quando $\lambda \rightarrow \infty$

$$\frac{X - \lambda}{\sqrt{\lambda}} \sim \mathcal{N}(0, 1).$$

Observação: Quando considerámos a aproximação da distribuição binomial pela Poisson, ambas eram discretas. Porém os dois teoremas acabados de enunciar dão-nos uma aproximação duma v.a discreta por uma contínua. Neste caso é necessário fazer-se o que se designa por **correção de continuidade** que consiste em considerar todo o inteiro k representado pelo intervalo $(k - 1/2, k + 1/2)$.

Exercício 2.16

Seja $X \sim B(15, 0.4)$ Calcule os valores exactos e os valores aproximados das seguintes probabilidades:

1. $P[X \leq 5]$;
2. $P[10 < X \leq 14]$;
3. $P[X = 12]$.

Resolução:

Como $X \sim \mathcal{N}(6, 1.9)$, então o cálculo do valor aproximado daquelas probabilidades utiliza a lei normal.

1. $P[X \leq 5] = 0.403$ (valor exacto).
 $P[X \leq 5] \sim P[X < 5.5] = \Phi\left(\frac{5.5-6}{1.9}\right) = 0.3974$ (valor aproximado).
2. $P[10 < X \leq 14] = 0.009$ (valor exacto).
 $P[10 < X \leq 14] \sim P[10.5 < X < 14.5] = \Phi(4.47) - \Phi(2.37) = 0.0089$ (valor aproximado).
3. $P[X = 12] = 0.0016$ (valor exacto).
 $P[X = 12] = P[11.5 < X < 12.5] = \Phi\left(\frac{12.5-6}{1.9}\right) - \Phi\left(\frac{11.5-6}{1.9}\right) = \Phi(3.42) - \Phi(2.89) = 0.0016$ (valor aproximado).

Muitos outros modelos contínuos surgem nas mais diversas áreas. Um dos que assume particular importância é o **modelo gama** e alguns dos seus casos particulares. A distribuição gama surge como um modelo possível para a caracterização de fenómenos que apresentem uma marcada assimetria.

Apresentamos de seguida a distribuição gama, apesar de não ter sido incluída no programa leccionado em 2007/2008. Consideramos, mais uma vez, que poderá vir a constituir material de apoio aos alunos em estudos futuros. A distribuição exponencial, que é um caso particular da distribuição gama, é estudada no presente ano lectivo. A distribuição qui-quadrado, aqui apresentada como caso particular da distribuição gama, só será tratada, em 2007/2008, no contexto da Inferência Estatística, como a distribuição de amostragem de uma variável função da variância amostral.

Distribuição gama e a distribuição exponencial

Embora a distribuição normal seja, como dissemos, um modelo de resolução de muitos problemas nas mais variadas áreas das ciências, há ainda muitas situações que requerem um tipo diferente de função densidade.

Começemos por referir brevemente a **distribuição gama** que deve o seu nome à **função gama**, estudada em muitas áreas da matemática.

Definição 2.21

A **função gama** é definida por

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx \quad \text{para } \alpha > 0. \quad (2.26)$$

Trata-se evidentemente de um integral convergente.

Vejamos algumas propriedades da função gama:

Propriedades

1. $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ (fórmula de recorrência da função gama)

Dem: Tem-se

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx ; \quad \text{integrando por partes vem}$$

$$[-e^{-x} x^{\alpha-1}]_0^{\infty} + \int_0^{\infty} e^{-x} (\alpha-1)x^{\alpha-2} dx = 0 + (\alpha-1) \int_0^{\infty} e^{-x} x^{\alpha-2} dx = (\alpha-1)\Gamma(\alpha-1).$$

2. No caso de $\alpha = n$ inteiro, é fácil verificar que se tem

$$\Gamma(n) = (n - 1)(n - 2)\dots\Gamma(1)$$

e como

$$\Gamma(1) = \int_0^{+\infty} e^{-x} dx = 1 \text{ vem}$$

$$\Gamma(n) = (n - 1)!$$

3. $\Gamma(1/2) = \sqrt{\pi}$

Dem: Deixa-se ficar como exercício a demonstração desta propriedade, tomando como sugestão que no integral que define $\Gamma(1/2)$ se faça a transformação $x = y^2/2$.

4. $\Gamma(\alpha)$ tem um único mínimo para $\alpha_0 = 1.4616$ sendo $\Gamma(\alpha_0) = 0.8856\dots$

Tem-se ainda $\lim_{\alpha \rightarrow 0} \Gamma(\alpha) = \lim_{\alpha \rightarrow \infty} \Gamma(\alpha) = +\infty$.

5. As derivadas da função gama são assim definidas:

$$\Gamma^{(k)}(\alpha) = \int_0^{\infty} x^{\alpha-1} (\log x)^k e^{-x} dx$$

Alguns valores particulares das derivadas úteis em muitas aplicações são

$\Gamma'(1) = \gamma = .57722\dots$ este valor é designado por **constante de Euler**

$\Gamma''(1) = \gamma^2 + \pi^2/6 = 1.97811\dots$

Vejamos agora como é definida a distribuição gama.

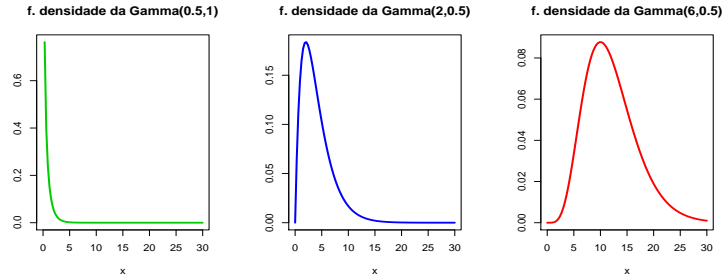
Uma v.a. diz-se ter **distribuição gama** de parâmetros α e β , ($\alpha > 0$, $\beta > 0$) e escreve-se $\mathbf{X} \sim \mathbf{G}(\alpha, \beta)$ se a função densidade é da forma

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (2.27)$$

Alguns gráficos desta função densidade, para vários valores de α e β , podem ver-se na figura seguinte.

Exercício 2.17

1. Verificar que $f(x)$ definida em (2.27) é de facto uma função densidade.



Gráficos da função densidade de uma v.a. com distribuição $G(1/2, 1)$, $G(2, 0.5)$ e $G(6, 0.5)$, da esquerda para a direita, respectivamente.

2. Provar que a função geradora de momentos da distribuição gama é

$$M_X(t) = \left(\frac{1}{1 - \beta t} \right)^\alpha \quad \text{para } t < 1/\beta.$$

3. Prove ainda que

$$E[X] = \alpha \beta \quad \text{Var}[X] = \alpha \beta^2.$$

Um caso particular muito importante é o que se obtém fazendo $\alpha = 1$.

A v.a. resultante diz-se ter **distribuição exponencial**⁴, representa-se por $\mathbf{X} \sim \mathbf{Exp}(\beta)$ e a função densidade é assim definida,

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & x > 0 \quad \beta > 0 \\ 0 & x \leq 0 \end{cases} \quad (2.28)$$

Como facilmente se pode verificar

A função geradora de momentos, o valor médio e variância da distribuição exponencial são assim definidos

$$M_X(t) = \frac{1}{1 - \beta t} \quad \text{para } t < 1/\beta \quad \text{e}$$

$$E[X] = \beta \quad \text{e} \quad \text{Var}[X] = \beta^2.$$

⁴Esta é a distribuição incluída no programa de 2007/2008

Exercício 2.18

Provar que se $X_i, i = 1, \dots, n$ são variáveis aleatórias independentes e semelhantes, com distribuição $Exp(\beta)$, então

$$\sum_{i=1}^n X_i \sim G(n, \beta).$$

Observações:

- A distribuição exponencial tem sido largamente usada como um modelo de problemas relativos a duração de vida, teoria da fiabilidade, tempos de espera, etc.
- Existe uma relação muito importante entre a distribuição exponencial e a distribuição de Poisson, que surge muitas vezes na prática. Suponhamos que estamos a observar a ocorrência de certos acontecimentos em intervalos de tempo. Seja T o tempo ao fim do qual se verifica a primeira ocorrência, trata-se de um variável aleatória contínua.

Teorema 2.18

Seja X uma v.a. de Poisson de parâmetro λ . Seja W a v.a. que designa o tempo de espera pela ocorrência do primeiro acontecimento, então W tem distribuição exponencial de parâmetro $\beta = 1/\lambda$.

A distribuição Qui-Quadrado

Um outro caso particular da distribuição gama, dá origem a uma distribuição muito importante, **a distribuição qui-quadrado**. Esta distribuição é largamente usada em inferência estatística, nomeadamente no estudo da variância de uma população a partir do estudo de uma amostra. No capítulo da Inferência Estatística irá ser largamente utilizada

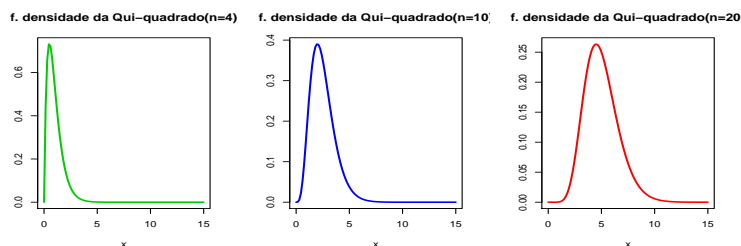
Consideremos uma v.a. com distribuição gama de parâmetros $\alpha = n/2$ (n inteiro positivo) e $\beta = 2$. Esta v.a. diz-se ter **distribuição qui-quadrado com n graus de liberdade** e representa-se por $X \sim \chi_{(n)}^2$.

A função densidade de X é

$$f(x) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} e^{-x/2} x^{n/2-1} & x > 0, n > 0 \\ 0 & x \leq 0. \end{cases} \quad (2.29)$$

Nota: Observe-se que a v.a. $X \sim \chi_{(n)}^2$, fica completamente especificada pela indicação do número de graus de liberdade n .

Vejamos alguns aspectos da densidade de uma v. a. com distribuição $\chi_{(n)}^2$, para alguns valores do parâmetro n .



Gráficos da função densidade de uma v.a. com distribuição $\chi^2_{(4)}$, $\chi^2_{(10)}$ e $\chi^2_{(20)}$, da esquerda para a direita, respectivamente.

Uma vez que se trata de um caso particular da distribuição gama é fácil obter

A Função geradora de momentos, valor médio e variância da distribuição

$$M_X(t) = \left(\frac{1}{1-2t} \right)^{n/2} \quad t < 1/2$$

$$E[X] = n \quad Var[X] = 2n \quad (2.30)$$

A distribuição Qui-quadrado encontra-se também tabelada, sendo a consulta das tabelas fornecidas para a disciplina de Estatística, feita do seguinte modo:

Se $X \sim \chi^2_{(n)}$, a cada par (n, α) corresponde um valor que designaremos por $\chi^2_{\alpha(n)}$ tal que $P(X > \chi^2_{\alpha(n)}) = \alpha$.

Exercício 2.19

Seja $X \sim \chi^2_{(15)}$ calcule

1. o ponto $\chi^2_{0.05}$;
2. o ponto A tal que $P(X < A) = 0.75$;

Resolução:

1. $\chi^2_{0.05} = 24.9958$;
2. $P(X < A) = 0.75 \Leftrightarrow P(X > A) = 0.25$, donde $A = 18.2451$.

Também esta distribuição goza da propriedade da estabilidade das somas. Vejamos o

Teorema 2.19

Sejam X_i , $i = 1, \dots, n$ variáveis aleatórias independentes tais que $X_i \sim \chi^2_{(n_i)}$. Então

$$\sum_{i=1}^n X_i \sim \chi^2_{(\sum n_i)}$$

Dem: A demonstração é imediata usando a f.g.m. da distribuição Qui-quadrado e a propriedade 2. da f.g.m. Deixa-se, por isso, como exercício.

Esta distribuição pode surgir porém, em estatística aplicada, definida de outro modo. **É desta forma que neste ano lectivo (2007/2008) ela será introduzida na Inferência Estatística.**

Vejam os seguinte teorema.

Teorema 2.20

Seja $Z \sim \mathcal{N}(0, 1)$ então a v.a. $X = Z^2$ tem distribuição $\chi^2_{(1)}$.

Dem:

Seja então $X = Z^2$ e vamos designar por $G(x)$ a função de distribuição de X . Tem-se então

$$\begin{aligned} G(x) &= P[X \leq x] = P[Z^2 \leq x] = P[-\sqrt{x} \leq Z \leq \sqrt{x}] = \\ &= \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) = 2\Phi(\sqrt{x}) - 1. \end{aligned}$$

Vejam agora qual a expressão da função densidade de X , que depois comparamos com a expressão (2.29) tomando $n = 1$.

$$\begin{aligned} g(x) &= G'(x) = \Phi'(\sqrt{x}) \cdot \frac{1}{\sqrt{x}} = \frac{1}{\sqrt{2\pi}} e^{-x/2} \frac{1}{\sqrt{x}} \\ &= \frac{1}{\sqrt{2}\sqrt{\pi}} e^{-x/2} x^{-1/2} = \frac{1}{\sqrt{2}\Gamma(1/2)} e^{-x/2} x^{-1/2} \end{aligned}$$

que de facto é a densidade de uma v.a. com distribuição $\chi^2_{(1)}$.

Generalização

Dadas Z_1, Z_2, \dots, Z_n , variáveis aleatórias independentes, normais reduzidas, a variável aleatória

$$X = Z_1^2 + Z_2^2 + \dots + Z_n^2$$

tem distribuição $\chi^2_{(n)}$.

Esta generalização é consequência imediata dos teoremas (2.19) e (2.20).

Exercícios propostos

1. Um automóvel tem instalado um sistema de alarme contra roubo. Sabe-se que, na área de estacionamento deste automóvel, a probabilidade de um automóvel ser assaltado é cerca de 5% e que a probabilidade do alarme disparar no caso de tentativa de roubo é de 99%.

A probabilidade de o alarme disparar sem ter havido tentativa de roubo é aproximadamente 2%.

- a) Calcule a probabilidade de:
- ter havido tentativa de roubo, sabendo que o alarme não disparou;
 - ter havido tentativa de roubo sabendo que o alarme disparou.
- b) Considere que na área a que se refere o problema estão estacionados 500 automóveis. Calcule, aproximadamente, a probabilidade de quando muito 10 destes automóveis serem assaltados.

2. Um posto de gasolina é reabastecido uma vez por semana. As vendas do passado sugerem que a função densidade de probabilidade do volume de vendas semanais, X , medido em dezenas de milhares de litros, é dada por:

$$f_X(x) = \begin{cases} x - 1 & \text{se } 1 \leq x \leq 2 \\ 3 - x & \text{se } 2 < x \leq 3 \\ 0 & \text{o.v.} \end{cases}$$

- a) Calcule a probabilidade de numa semana o volume de vendas se situar entre os 1500 litros e os 2300 litros.
- b) Calcule o valor esperado e o desvio padrão do volume de vendas semanais.
- c) Determine a quantia mínima com que o posto se deve abastecer, por semana, para que o camião tanque abastecedor não encontre a gasolina esgotada no posto em mais de 8% das semanas.
- d) Admitindo que o volume de vendas é independente de semana para semana, calcule a probabilidade de em 2 anos o posto vender mais de 210 dezenas de milhares de litros.
- e) O referido posto de gasolina situa-se numa via ao longo da qual os postos se distribuem segundo uma lei de Poisson com média de 1 posto por 10 km. Ocorreu uma greve no sistema de abastecimento de cada posto. Devido a isto cada posto tem, independentemente dos outros, uma probabilidade 0.2 de estar esgotado.
- i) Qual a probabilidade de não existirem mais de dois postos nos próximos 30 km?

- ii) Qual a probabilidade de nos próximos 3 postos nenhum ter gasolina para vender?
3. A distribuição conjunta de X , proporção da capacidade do depósito de um automóvel que é reabastecida no início de cada semana e Y a proporção da capacidade gasta durante a semana é dada por:

$$f(x, y) = \begin{cases} 3x & \text{se } 0 < y < x < 1 \\ 0 & \text{o.v. de } (x, y) \end{cases}$$

- a) Calcule as funções densidade marginais de X e Y . Serão X e Y variáveis aleatórias independentes? Justifique.
- b) Calcule a probabilidade de a quantidade gasta ser inferior a metade da reabastecida.
- c) Qual é a diferença média entre a proporção de reabastecimento e de gasto?
4. O tempo gasto na tosquia de uma ovelha por um tosquiador experiente considera-se uma v.a. com distribuição normal de média 10 min e desvio padrão 2 min. Determine a probabilidade de:
- a) a tosquia de uma ovelha, escolhida ao acaso, demorar mais do que 13 minutos;
- b) pelo menos uma de 4 ovelhas, escolhidas ao acaso, necessitar de mais de 13 minutos para ser tosquiada (admita o tempo de tosquia independente de ovelha para ovelha);
- c) o tempo total gasto na tosquia de um rebanho, constituído por 20 ovelhas, ser inferior a 3 horas (admita o tempo de tosquia independente de ovelha para ovelha).
5. O tempo de vida Y de determinado equipamento é dado por “ $Y = 500 + X$ ” horas, onde X é uma variável aleatória exponencial de parâmetro $\beta = 40$.
- a) Determine o tempo médio de vida deste equipamento.
- b) Qual a probabilidade de que o equipamento dure pelo menos 580 horas?
- c) Dados 100 equipamentos do tipo anterior, a funcionar independentemente, determine, justificando, o valor aproximado da probabilidade de que no máximo três durem 580 horas ou mais.

Referências bibliográficas

Bhattacharyya, G.K. and Johnson R.A.(1977), *Statistical Concepts and Methods*, John Wiley & Sons Inc.

- Dagnelie, P.(1973), *Estatística, Teoria e Métodos*, trad. do Prof. Doutor A. St.Aubyn, Europa América, vol I e II.
- Dagnelie, P.(1998), *Statistique Théorique et Appliqué*, Tome 1, De Boeck Université.
- Daniel, W. W.(1995), *Biostatistics: A Foundation for Analysis in the Health Sciences*. John Wiley.
- Milton, J.S. and Arnold, J.C. (1987), *Probability and Statistics in the Engineering and Computing Sciences*, Mc Graw Hill.
- Montgomery, D.C. e Runger, G.C.(1994) - *Applied Statistics and Probability for Engineers*. John Wiley.
- Mood, A. Graybill, F. and Boes D. (1985). *Introduction to the Theory of Statistics* , Mc Graw Hill.
- Murteira, B. (1990), *Probabilidades e Estatística* (vol I), Mc Graw Hill.
- Murteira, B., Ribeiro, C.S., Silva, J.A. e Pimenta C.(2002), *Introdução à Estatística*, Mc Graw Hill.
- Pestana, D.D. e Velosa, S.F. (2002), *Introdução à Probabilidade e à Estatística* . Fundação Calouste Gulbenkian.
- Tiago de Oliveira, J. (1990), *Probabilidades e Estatística. Conceitos, Métodos e Aplicações* (vol I), Mc Graw Hill.
- Walpole, R.E (1993), *Mathematical Statistics*, 3th edition, Englewood Cliffs, N.J., Prentice-Hall.