

Capítulo II: Variáveis Aleatórias

Elementos de Probabilidades e Teoria de Números

Licenciatura em Engenharia Informática
e
Mestrado Integrado em Engenharia Informática

Universidade do Minho
Ano Letivo 2021/2022

Neste capítulo pretende-se apresentar alguns dos modelos aleatórios (ou estocásticos) mais utilizados em aplicações a problemas reais. Em particular, iremos:

- apresentar o conceito de variável aleatória;
- distinguir entre variáveis aleatórias discretas e variáveis aleatórias contínuas;
- apresentar funções que “caracterizam” as variáveis aleatórias (e.g., função massa de probabilidade, função densidade de probabilidade e função de distribuição).
- apresentar algumas medidas de localização e de dispersão de uma variável aleatória (valor médio, desvio-padrão/variância, quantis);
- conhecer algumas das distribuições de probabilidade mais usadas em modelação estocástica e/ou em inferência estatística (e.g., Binomial, Poisson, Exponencial, Normal, etc.)

1. Variáveis aleatórias

Em muitas situações práticas interessa-nos (ou é conveniente) associar valores numéricos aos diferentes resultados de uma experiência aleatória, ou seja, a cada elemento ω do espaço de resultados Ω interessa associar um número real. Esta associação permitirá passar do espaço amostral Ω , por vezes complicado, para um conjunto mais simples (\mathbb{R} ou um seu subconjunto).

Por exemplo, na experiência aleatória que consiste em efectuar vários lançamentos consecutivos de um dado podemos estar interessados em estudar

- o valor da soma das faces obtidas;
- o número de ases obtidos;
- o número de faces par obtidas;
- o máximo (ou o mínimo) das faces obtidas,

etc.

Nota: No lançamento de um dado, “sair ás” corresponde a “sair face 1”.

1. Variáveis aleatórias

Surge assim a definição de *variável aleatória*.

Definição

Seja Ω o espaço amostral associado a uma experiência aleatória. Uma variável aleatória (v.a.) é uma função cujo domínio é o conjunto Ω e cujo conjunto de chegada é \mathbb{R} , que a cada elemento ω de Ω faz corresponder um número real, i.e.,

$$\begin{array}{ccc} X : & \Omega & \longrightarrow & \mathbb{R} \\ & \omega & \longmapsto & X(\omega) \end{array} .$$

Notas:

- 1) As v.a.'s são habitualmente denotadas com letras maiúsculas, e.g., X , Y , Z , etc.
- 2) As variáveis aqui estudadas são conhecidas por v.a.'s reais (devido ao conjunto de chegada). Existem, no entanto, outro tipo de variáveis como, por exemplo, v.a.'s complexas ($X : \Omega \rightarrow \mathbb{C}$).

1. Variáveis aleatórias

Exemplo: Na experiência aleatória que consiste em efectuar dois lançamentos consecutivos de uma moeda, o espaço amostral é, como bem sabemos,

$$\Omega = \{(Ca, Ca), (Ca, Co), (Co, Ca), (Co, Co)\},$$

onde Ca e Co representam cara e coroa, respectivamente.

Consideremos agora as v.a.'s X e Y que representam, respectivamente, o **número de coroas** e a **diferença entre o número de caras e o número de coroas**. As funções X e Y estão descritas na seguinte tabela:

ω	$X(\omega)$	$Y(\omega)$
(Ca,Ca)	0	2
(Ca,Co)	1	0
(Co,Ca)	1	0
(Co,Co)	2	-2

1. Variáveis aleatórias

Dado um subconjunto $B \subseteq \mathbb{R}$, a probabilidade de a v.a. X assumir um valor pertencente a B , i.e. $P(X \in B)$, será dada pela probabilidade do subconjunto formado pelos elementos de Ω cuja imagem, pela função X , pertence a B .

Em linguagem matemática, o subconjunto formado pelos elementos de Ω cuja imagem, pela função X , pertence a B é designado por *imagem inversa de B pela função X* e é denotado por $X^{-1}(B)$:

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

Assim, depois de identificado $X^{-1}(B)$, teremos simplesmente

$$P(X \in B) = P(X^{-1}(B)).$$

1. Variáveis aleatórias

Voltando ao exemplo dos dois lançamentos de uma moeda:

O acontecimento “saiu pelo menos uma coroa” é agora representado por $(X \geq 1)$ ou, de modo equivalente, $(X \in [1, +\infty[)$.

O acontecimento “número de caras igual ao número de coroas” é representado por $(Y = 0)$ ou, de modo equivalente, $(Y \in \{0\})$.

Para calcular $P(X \in [1, +\infty[)$ e $P(Y \in \{0\})$ temos primeiro que identificar os seguintes subconjuntos de Ω :

$$X^{-1}([1, +\infty[) \quad \text{e} \quad Y^{-1}(\{0\}).$$

Observando a tabela atrás vemos facilmente que

$$X^{-1}([1, +\infty[) = \{(Ca, Co), (Co, Ca), (Co, Co)\}$$

e que

$$Y^{-1}(\{0\}) = \{(Ca, Co), (Co, Ca)\}.$$

1. Variáveis aleatórias

Se a moeda for equilibrada, os acontecimentos elementares são equiprováveis, i.e.,

$$P(\{(Ca, Ca)\}) = P(\{(Ca, Co)\}) = P(\{(Co, Ca)\}) = P(\{(Co, Co)\}) = \frac{1}{4}.$$

Podemos então usar a definição de Laplace para calcular as probabilidades dos diferentes acontecimentos decorrentes desta experiência. Assim,

$$P(X \geq 1) = P(X^{-1}([1, +\infty[)) = P(\{(Ca, Co), (Co, Ca), (Co, Co)\}) = \frac{3}{4}$$

e

$$P(Y = 0) = P(Y^{-1}(\{0\})) = P(\{(Ca, Co), (Co, Ca)\}) = \frac{2}{4} = \frac{1}{2}.$$

1. Variáveis aleatórias

Entre as v.a.'s distinguem-se v.a.'s discretas e v.a.'s contínuas.

Uma v.a. X diz-se *discreta* se o seu contradomínio (i.e., o conjunto de valores que a função X assume) é um conjunto finito (tem cardinal finito) ou infinito numerável (e.g., \mathbb{N} , \mathbb{N}_0 , \mathbb{Z} , \mathbb{Z}^+ , \mathbb{Z}^-).

Note-se que se o espaço amostral Ω for finito ou infinito numerável, então qualquer v.a. sobre ele definido será discreta (pela definição de função, o número de elementos do contradomínio nunca poderá ser superior ao número de elementos do domínio Ω).

Mas, se Ω for infinito não numerável (e.g., \mathbb{R} , \mathbb{R}^-), podemos definir v.a.'s discretas mas também não discretas.

2. Variáveis aleatórias discretas

Uma v.a. discreta, X , é caracterizada pelo seu contradomínio, denotado por C_X , que é o conjunto de valores que a v.a. assume com probabilidade estritamente positiva, e ainda por uma função, designada de *função massa de probabilidade* (f.m.p.), que a cada elemento $a \in C_X$ faz corresponder o valor $P(X = a)$.

Sendo X uma v.a. discreta com contradomínio $C_X = \{x_1, x_2, x_3, \dots\}$, a sua f.m.p. é usualmente representada do seguinte modo:

$$X: \begin{array}{cccc} x_1 & x_2 & x_3 & \dots \\ P(X = x_1) & P(X = x_2) & P(X = x_3) & \dots \end{array}.$$

No exemplo dos dois lançamentos da moeda, as f.m.p.'s de X e Y são, respetivamente:

$$X: \begin{array}{ccc} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{array} \quad \text{e} \quad Y: \begin{array}{ccc} -2 & 0 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{array}.$$

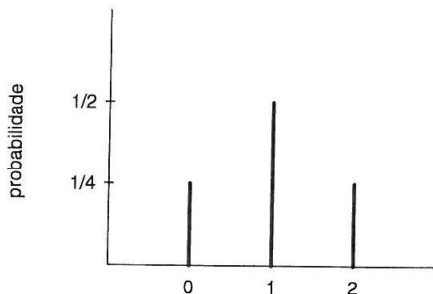
2. Variáveis aleatórias discretas

Graficamente, a f.m.p. representa-se através de um diagrama de linhas, semelhante ao diagrama de barras que usamos para representar as frequências relativas simples de uma amostra proveniente de uma variável quantitativa discreta.

De facto, a f.m.p. é a versão teórica - ou populacional- que tem a correspondente versão amostral nos diagramas de barras utilizados em estatística descritiva para a representação de dados discretos.

2. Variáveis aleatórias discretas

A representação gráfica da f.m.p. da variável X é a seguinte:



2. Variáveis aleatórias discretas

A partir da definição axiomática de probabilidade e da definição de v.a. deduz-se o seguinte: se X é uma v.a. discreta, com contradomínio C_X , então

$$\sum_{x_i \in C_X} P(X = x_i) = 1.$$

Deduz-se também a seguinte fórmula geral para o cálculo de probabilidades envolvendo uma v.a. X discreta: para um qualquer subconjunto $B \subseteq \mathbb{R}$, tem-se

$$P(X \in B) = \sum_{x_i \in C_X : x_i \in B} P(X = x_i).$$

No exemplo dos dois lançamentos da moeda temos, para a v.a. X ,

$$P(0 \leq X \leq 1.5) = \sum_{x_i \in C_X : x_i \in [0, 1.5]} P(X = x_i) = P(X = 0) + P(X = 1) = \frac{3}{4}$$

e, para a v.a. Y ,

$$P(Y > 0) = \sum_{y_i \in C_Y : y_i \in]0, +\infty[} P(Y = y_i) = P(Y = 2) = \frac{1}{4}.$$

3. Variáveis aleatórias contínuas

Entre as v.a.'s que têm contradomínio infinito não numerável (e.g., \mathbb{R}, \mathbb{R}^- , $[0, +\infty[$, um qualquer intervalo real de amplitude não nula), vamos estudar as chamadas v.a.'s *absolutamente contínuas* que são caracterizadas à custa de uma função designada de *função densidade de probabilidade*.

Definição

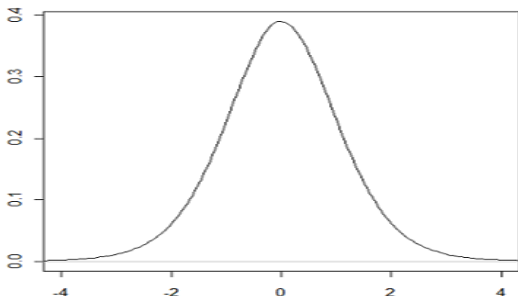
Uma função $f : \mathbb{R} \rightarrow \mathbb{R}$ diz-se uma função densidade de probabilidade se satisfaz as seguintes condições:

- i) para todo o $x \in \mathbb{R}$, $f(x) \geq 0$,
- ii) $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Nota: Nesta disciplina, as v.a.'s *absolutamente contínuas* serão simplesmente designadas de *contínuas*.

3. Variáveis aleatórias contínuas

De seguida representa-se, a título de exemplo, o gráfico de uma função densidade de probabilidade: uma função não negativa e tal que a área da região compreendida entre o eixo dos $xx's$ e o gráfico da função é igual a 1.



Note-se que a função densidade de probabilidade é a versão teórica - ou populacional - do histograma das frequências relativas usado em estatística descritiva para representar dados contínuos (recordar que a área total de tal histograma é igual a 1).

3. Variáveis aleatórias contínuas

Para uma v.a. contínua, X , com função densidade de probabilidade f , tem-se que: para um qualquer subconjunto $B \subseteq \mathbb{R}$,

$$P(X \in B) = \int_B f(x)dx.$$

Note que, se B é um conjunto singular, i.e. $B = \{b\}$, com b um número real, tem-se

$$P(X \in B) = P(X = b) = 0.$$

No caso em que B é um intervalo real da forma $[a, c]$, $[a, c[$, $]a, c]$ ou $]a, c[$, com $a < c$, tem-se

$$P(X \in B) = \int_a^c f(x)dx.$$

4. Função de distribuição de uma variável aleatória

Já vimos que as v.a.'s discretas são caracterizadas à custa da função massa de probabilidade e que as v.a.'s contínuas são caracterizadas à custa da função densidade de probabilidade. No entanto, **qualquer** v.a. é também caracterizada pela chamada *função de distribuição*.

Definição

Chamamos função de distribuição da v.a. X à função

$$\begin{aligned} F : \mathbb{R} &\longrightarrow [0, 1] \\ c &\mapsto F(c) = P(X \leq c) \end{aligned}$$

De entre as propriedades de uma função de distribuição destacam-se:

- i) F é uma função não decrescente,
- ii) F é contínua à direita,
- iii)

$$\lim_{c \rightarrow -\infty} F(c) = 0 \quad \text{e} \quad \lim_{c \rightarrow +\infty} F(c) = 1.$$

4. Função de distribuição de uma variável aleatória

Para uma v.a. discreta X , a função de distribuição obtém-se do seguinte modo

$$F(c) = P(X \leq c) = \sum_{x_i \in C_X : x_i \leq c} P(X = x_i).$$

Exemplo: Para a v.a. X que representa o número de coroas obtidas em dois lançamentos consecutivos de uma moeda equilibrada, a respetiva função de distribuição é dada por:

$$F(c) = P(X \leq c) = \begin{cases} 0 & \text{se } c < 0 \\ P(X = 0) = \frac{1}{4} & \text{se } 0 \leq c < 1 \\ P(X = 0) + P(X = 1) = \frac{3}{4} & \text{se } 1 \leq c < 2 \\ P(X = 0) + P(X = 1) + P(X = 2) = 1 & \text{se } c \geq 2 \end{cases}$$

Nota: Esboce o gráfico desta função e constate que esta verifica as propriedades i) a iii) referidas no slide anterior.

4. Função de distribuição de uma variável aleatória

Para uma v.a. contínua X , com função densidade de probabilidade f , a função de distribuição de X obtém-se do seguinte modo

$$c \in \mathbb{R}, F(c) = P(X \leq c) = \int_{-\infty}^c f(x)dx.$$

Exemplo: Para uma v.a. contínua X que tenha como função densidade de probabilidade

$$f(x) = \begin{cases} 10 & \text{se } -0.1 \leq x \leq 0 \\ 0 & \text{se } c.c. \end{cases},$$

a função de distribuição é dada por:

$$\begin{aligned} F(c) &= P(X \leq c) = \int_{-\infty}^c f(x)dx = \\ &= \begin{cases} \int_{-\infty}^c 0 dx = 0 & \text{se } c < -0.1 \\ \int_{-\infty}^{-0.1} 0 dx + \int_{-0.1}^c 10 dx = 10(c + 0.1) & \text{se } -0.1 \leq c < 0 \\ \int_{-\infty}^{-0.1} 0 dx + \int_{-0.1}^0 10 dx + \int_0^c 0 dx = 1 & \text{se } c \geq 0 \end{cases} \end{aligned}$$

4. Função de distribuição de uma variável aleatória

A função de distribuição é muito útil para calcular probabilidades do tipo

$$P(a < X \leq b)$$

com $a < b$ constantes reais. De facto,

$$\begin{aligned} P(a < X \leq b) &= P((X \leq b) \cap (X > a)) \\ &= P((X \leq b) \cap (\overline{X \leq a})) \\ &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a) \end{aligned}$$

5. Medidas de localização e de dispersão de uma v.a.

As v.a.'s têm características, designadas de *teóricas* ou *populacionais*, correspondentes às características amostrais calculadas em estatística descritiva (por exemplo: média, variância, desvio-padrão, mediana, percentis, etc.) para uma amostra dessas mesmas variáveis, e que pretendem quantificar determinados aspectos da sua distribuição (localização, dispersão/variabilidade, etc).

Definição

O *valor médio* de uma v.a. X (ou valor médio populacional), usualmente denotado por μ_X ou $E[X]$, é a média pesada, de acordo com a função massa/densidade de probabilidade, dos valores de X , i.e.,

$$\begin{array}{ll} \mu_X = \sum_{x_i \in C_X} x_i P(X = x_i) & , \quad \mu_X = \int_{-\infty}^{+\infty} x f(x) dx. \\ \text{(caso discreto)} & \text{(caso contínuo)} \end{array}$$

Obs.: A notação $E[X]$ deriva de "esperança matemática" que é uma designação alternativa para valor médio.

5. Medidas de localização e de dispersão de uma v.a.

Outra característica importante de uma v.a. X é a *variância*, usualmente denotada por σ_X^2 ou $Var[X]$.

Definição

Seja X uma v.a.. A *variância* de X é dada por

$$\sigma_X^2 = E[(X - \mu_X)^2]$$

e, tal como acontece com o valor médio, é calculada de forma diferente conforme X é discreta ou contínua. Assim, a variância de X é dada por:

$$\sigma_X^2 = \sum_{x_i \in C_X} (x_i - \mu_X)^2 P(X = x_i) \quad , \quad \sigma_X^2 = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f(x) dx.$$

(caso discreto) (caso contínuo)

A $\sigma_X = \sqrt{\sigma_X^2}$ chamamos *desvio-padrão* da v.a. X .

Obs.: Repare-se na semelhança destas fórmulas com as da média, variância e desvio-padrão amostrais usadas em estatística descritiva. ↻ 🔍

5. Medidas de localização e de dispersão de uma v.a.

Notas:

1) Recordar que as somas/integrais podem não existir ou não ser finitas. Nesse caso diz-se que a v.a. não tem a característica em causa.

2) Quando $E[X^2]$ existe, i.e., quando

$$\begin{array}{ll} \sum_{x_i \in C_X} x_i^2 P(X = x_i) < +\infty & , \quad \int_{-\infty}^{+\infty} x^2 f(x) dx < +\infty \\ \text{(caso discreto)} & \text{(caso contínuo)} \end{array}$$

a fórmula da variância reduz-se facilmente a

$$\sigma_X^2 = E[X^2] - (E[X])^2,$$

ou seja,

$$\sigma_X^2 = \left[\sum_{x_i \in C_X} x_i^2 P(X = x_i) \right] - (E[X])^2 \quad , \quad \sigma_X^2 = \left[\int_{-\infty}^{+\infty} x^2 f(x) dx \right] - (E[X])^2 .$$

(caso discreto)
(caso continuo)

3) Quando $E[X^k]$ existe, o seu valor é designado de *momento de ordem k de X* .

5. Medidas de localização e de dispersão de uma v.a.

Finalmente, os quantis teóricos, que desempenham um papel muito importante em inferência estatística, são definidos de modo análogo aos quantis amostrais.

Recordar que, em estatística descritiva, o quantil amostral de ordem p , é, grosso modo, o valor que separa os $100 \times p\%$ valores inferiores da amostra dos $100 \times (1 - p)\%$ superiores.

Definição

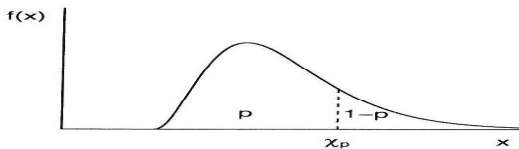
Para uma v.a. X , o *quantil de ordem p* , com $p \in]0, 1[$, denota-se por χ_p , é dado por:

$$\chi_p = \inf\{c \in \mathbb{R} : F(c) \geq p\},$$

onde F é a função de distribuição de X .

5. Medidas de localização e de dispersão de uma v.a.

Para uma v.a. contínua, o quantil χ_p é um valor que, no gráfico da função densidade de probabilidade, tem à sua esquerda área igual a p e à sua direita área igual a $(1 - p)$.



Assim, neste caso e se $F^{-1}(p)$ existir, tem-se que

$$\chi_p = F^{-1}(p).$$

Casos especiais de quantis:

- 1 Mediana: $\chi_{0.5}$
- 2 Quartis: $\chi_{0.25}, \chi_{0.5}, \chi_{0.75}$
- 3 Decis: $\chi_{0.1}, \chi_{0.2}, \chi_{0.3}, \chi_{0.4}, \chi_{0.5}, \chi_{0.6}, \chi_{0.7}, \chi_{0.8}, \chi_{0.9}$
- 4 Percentis: $\chi_{0.01}, \chi_{0.02}, \dots, \chi_{0.98}, \chi_{0.99}$

6. Independência de variáveis aleatórias

No Capítulo I, vimos a definição de acontecimentos independentes. Na altura ficou subjacente a ideia de que a independência entre dois acontecimentos se baseava no facto de o conhecimento da realização de um deles não afetar a probabilidade de ocorrência do outro (fica ao cuidado do aluno estender esta ideia a mais do que 2 acontecimentos). Vejamos agora a definição de variáveis aleatórias independentes.

Definição

Dadas $n \geq 2$ variáveis aleatórias, X_1, X_2, \dots, X_n , estas dizem-se independentes se, para todos os B_1, B_2, \dots, B_n subconjuntos de \mathbb{R} ,

$$P((X_1, X_2, \dots, X_n) \in (B_1 \times B_2 \times \dots \times B_n)) = \prod_{i=1}^n P(X_i \in B_i) \quad (1)$$

Notas:

1) Observe que a condição (1) pode ser escrita do seguinte modo:

$$P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = P(X_1 \in B_1)P(X_2 \in B_2) \dots P(X_n \in B_n).$$

2) Entre v.a.'s a “,” substituí a “ \cap ”, pelo que

$$P(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) = P((X_1 \in B_1) \cap (X_2 \in B_2) \cap \dots \cap (X_n \in B_n))$$

6. Independência de variáveis aleatórias

Regra geral, é difícil provar que variáveis aleatórias são independentes. Existem resultados importantes que nos permitem averiguar, com mais facilidade, sobre a independência. Vamos enunciar aqui alguns.

Teorema

Sejam X_1, X_2, \dots, X_n variáveis aleatórias.

- i) X_1, X_2, \dots, X_n são independentes sse, para quaisquer números reais c_1, c_2, \dots, c_n , se tem

$$P(X_1 \leq c_1, X_2 \leq c_2, \dots, X_n \leq c_n) = \prod_{i=1}^n P(X_i \leq c_i).$$

- ii) Se X_1, X_2, \dots, X_n são v.a.'s discretas, então X_1, X_2, \dots, X_n são independentes sse, para quaisquer números reais a_1, a_2, \dots, a_n , se tem

$$P(X_1 = a_1, X_2 = a_2, \dots, X_n = a_n) = \prod_{i=1}^n P(X_i = a_i).$$

- iii) Se X_1, X_2, \dots, X_n são independentes e $\phi_1, \phi_2, \dots, \phi_n$ são funções reais de variável real tais que $\phi_i(X_i)$, $i = 1, \dots, n$, ainda são v.a.'s, então $\phi_1(X_1), \phi_2(X_2), \dots, \phi_n(X_n)$ também são v.a.'s independentes.

6. Independência de variáveis aleatórias

Vamos terminar esta secção enunciando algumas propriedades sobre o valor médio e a variância de transformações lineares, somas e produtos de v.a.'s (discretas ou contínuas). Algumas destas propriedades dizem respeito a v.a.'s independentes.

- 1) Para quaisquer números reais, a e b , e para qualquer v.a. X , tem-se

$$E[aX + b] = aE[X] + b \text{ e } Var[aX + b] = a^2 Var[X].$$

- 2) Para quaisquer v.a.'s X_1, X_2, \dots, X_n , tem-se

$$E[X_1 + X_2 + \dots + X_n] = E[X_1] + E[X_2] + \dots + E[X_n].$$

- 3) Se X_1, X_2, \dots, X_n são v.a.'s independentes então

$$Var[X_1 + X_2 + \dots + X_n] = Var[X_1] + Var[X_2] + \dots + Var[X_n].$$

- 4) Se X_1, X_2, \dots, X_n são v.a.'s independentes então

$$E[X_1 X_2 \dots X_n] = E[X_1]E[X_2] \dots E[X_n].$$

7. Algumas distribuições de probabilidade

Nesta secção pretende-se apresentar algumas distribuições de probabilidade mais conhecidas e mais utilizadas na prática, em problemas de modelação estocástica e/ou de inferência estatística. Exemplos de algumas dessas distribuições são:

- Binomial (e Bernoulli, como um caso particular)
- Poisson
- Geométrica
- Uniforme
- Exponencial
- Normal

7.1. Distribuição Binomial, com parâmetros n e p

Definição

Considere que, numa experiência aleatória ξ , um acontecimento S ocorre com probabilidade p , com $p \in]0, 1[$. Considere agora uma v.a. X que representa o número de vezes que S ocorre em n repetições independentes da experiência ξ . Tem-se que X é uma v.a. discreta, com contradomínio $C_X = \{0, 1, \dots, n\}$, e a sua função massa de probabilidade é dada por

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

Nestas condições, diz-se que a v.a. X segue a distribuição (ou lei) Binomial com parâmetros n e p , e abrevia-se por $X \sim \text{Bin}(n, p)$.

Nota: O acontecimento S é usualmente designado de "sucesso".

7.1. Distribuição Binomial, com parâmetros n e p

Observações:

1) Se $X \sim \text{Bin}(n, p)$, então o valor médio e a variância são

$$E[X] = np \quad \text{e} \quad \text{Var}[X] = np(1 - p).$$

2) A distribuição $\text{Bin}(n, p)$ também é a distribuição de uma v.a. X definida nas seguintes condições: *Suponha que numa população existe uma proporção $p \in]0, 1[$ de indivíduos que tem uma certa característica A (e uma proporção $1 - p$ não tem a característica). Considere agora a experiência aleatória que consiste em escolher ao acaso, e **com reposição**, n indivíduos desta população e seja X a v.a. que representa o número de indivíduos escolhidos que possuem a característica A . Então $X \sim \text{Bin}(n, p)$.*

3) Quando $n = 1$, a distribuição binomial é conhecida por *Bernoulli*(p). Neste caso, o contradomínio é $C_X = \{0, 1\}$ e a f.m.p. é dada por

$$X : \begin{cases} 0 & 1 \\ 1 - p & p \end{cases}.$$

4) Se $X \sim \text{Bin}(n, p)$ então $X \stackrel{d}{=} \sum_{i=1}^n X_i$, em que X_1, X_2, \dots, X_n são v.a.'s independentes e identicamente distribuídas (i.i.d.'s) com lei *Bernoulli*(p).

7.1. Distribuição Binomial, com parâmetros n e p

Exemplos/Exercícios [Folha 4]: Em uma das seguintes alíneas, defina uma v.a. com distribuição Binomial que seja relevante para o problema.

- a) Sabendo que 60% dos indivíduos de uma determinada população são pobres, determine a probabilidade de numa amostra de 10 indivíduos escolhidos, ao acaso e com reposição, nesta população, haver exatamente 9 pobres? E de haver pelo menos 9 pobres? E de não haver qualquer pobre?
- b) Em 10 lançamentos consecutivos de uma moeda equilibrada, qual a probabilidade de saírem 2 caras? E qual a probabilidade de saírem pelo menos 2 caras?
- c) Uma urna contém 3 bolas brancas e 2 bolas vermelhas. Determine a probabilidade de ao extrair, com reposição, 4 bolas desta urna, todas as bolas escolhidas serem brancas. E qual a probabilidade de todas as bolas extraídas serem vermelhas?
E se a extracção for feita sem reposição?

Nota: Em relação a c), observe que, quando a escolha feita sem reposição, as diferentes extracções não podem ser vistas como independentes.

7.2. Distribuição de Poisson, com parâmetro λ

Historicamente, a distribuição de Poisson surge como sendo a distribuição limite de uma Binomial, em que os parâmetros n e p satisfazem certas condições. Vejamos então como surgiu esta distribuição.

Seja $(X_n)_{n \in \mathbb{N}}$ uma sucessão de v.a.'s tais que $X_n \sim \text{Bin}(n, p_n)$ e em que os parâmetros n e p_n satisfazem a seguinte condição

$$\lim_{n \rightarrow +\infty} np_n = \lambda,$$

com $\lambda \in \mathbb{R}^+$. Note-se que, nestas condições, tem-se

$$\lim_{n \rightarrow +\infty} p_n = 0$$

e, para todo o $k \in \mathbb{N}$,

$$\frac{P(X_n = k)}{P(X_n = k-1)} = \frac{\binom{n}{k} p_n^k (1-p_n)^{n-k}}{\binom{n}{k-1} p_n^{k-1} (1-p_n)^{n-k+1}} = \frac{n-k+1}{k} \frac{p_n}{1-p_n} \xrightarrow{n \rightarrow +\infty} \frac{\lambda}{k}.$$

Isto permite-nos concluir que, quando n é grande, a função massa de probabilidade da v.a. X_n comporta-se como a de uma v.a. discreta, Y , de contradomínio \mathbb{N}_0 , e tal que

$$P(Y = k) = \frac{\lambda}{k} P(Y = k-1), \quad k \in \mathbb{N}.$$

7.2. Distribuição de Poisson, com parâmetro λ

Trabalhando esta última igualdade, concluímos que

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

Como $C_Y = \mathbb{N}_0$, temos

$$1 = \sum_{k=0}^{+\infty} P(Y = k) \Leftrightarrow 1 = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} P(Y = 0) \underset{(*)}{\Leftrightarrow} 1 = P(Y = 0) e^{\lambda} \Leftrightarrow P(Y = 0) = e^{-\lambda}.$$

Concluimos assim que a função massa de probabilidade de Y é

$$P(Y = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \in \mathbb{N}_0. \quad (2)$$

Definição

Quando uma v.a. Y , discreta e de contradomínio \mathbb{N}_0 , tem função massa de probabilidade dada por (2), diz-se que Y segue uma *distribuição Poisson com parâmetro λ* , e abrevia-se por $Y \sim \text{Poisson}(\lambda)$.

7.2. Distribuição de Poisson, com parâmetro λ

Observações:

1) Na prática, a distribuição de Poisson é adequada para modelar o número de ocorrências de um fenómeno raro (i.e. um fenómeno que tem baixa probabilidade de ocorrência) quando não limitamos o número de repetições da experiência aleatória.

Em particular, a função massa de probabilidade da distribuição de Poisson é usada para obter um valor aproximado da função massa de probabilidade de uma v.a. $Z \sim \text{Bin}(n, p)$ quando n é grande e p é pequeno. O parâmetro λ a utilizar na aproximação será igual a $n \times p$.

2) Se $Y \sim \text{Poisson}(\lambda)$ então o valor médio e a variância são

$$E[Y] = \lambda \quad \text{e} \quad \text{Var}[Y] = \lambda,$$

respetivamente.

3) Da expansão em série de Taylor da função e^x , em torno de zero, resulta a seguinte igualdade usada em (*):

$$e^x = \sum_{n=0}^{+\infty} \frac{x^n}{n!}, \quad x \in \mathbb{R}.$$

7.2. Distribuição de Poisson, com parâmetro λ

Exemplo/Exercício: É editado um manual de probabilidades com uma tiragem de 100000 exemplares. A probabilidade de um manual ter defeito na encadernação é de 10^{-4} . Calcule a probabilidade de nesta tiragem de 100000 exemplares

- i) haver exactamente 5 manuais com defeito;
- ii) haver, no máximo, 2 manuais com defeito;
- iii) haver pelo menos 3 manuais com defeito.

Indique o valor exacto (recorrendo à distribuição Binomial) e o valor aproximado (recorrendo à distribuição de Poisson) de cada uma das probabilidades pedidas.

Solução: Valor exacto / valor aproximado:

- i) 0.03782949 / 0.03783327;
- ii) 0.002768488 / 0.002769396;
- iii) 0.9972315 / 0.9972306.

Nota: O valor exato foi obtido usando o software estatístico R.

7.3. Distribuição Geométrica, com parâmetro p

Considere uma experiência aleatória na qual um certo acontecimento, que designamos por “sucesso”, ocorre com probabilidade $0 < p < 1$ (e ocorre “insucesso” com probabilidade $1 - p$). Suponhamos agora que se repete esta experiência, sempre nas mesmas condições, e seja T a v.a. que representa o número de vezes que se efetua a experiência até ocorrer “sucesso” pela primeira vez.

Naturalmente, tem-se que T é discreta e que $C_T = \{1, 2, \dots\} = \mathbb{N}$. Para determinar a função massa de probabilidade de T , considerem-se os seguintes acontecimentos:

A_i : “ocorreu insucesso na i -ésima vez que se efetuou a experiência”, $i = 1, 2, \dots$. Note-se que $P(A_i) = 1 - p$ e $P(\overline{A_i}) = p$, para todo o i .

Usando estes acontecimentos, podemos calcular alguns dos valores da função massa de probabilidade de T :

$$P(T = 1) = P(\overline{A_1}) = p;$$

$$P(T = 2) = P(A_1 \cap \overline{A_2}) = P(A_1) \times P(\overline{A_2}) = (1 - p)p;$$

$$P(T = 3) = P(A_1 \cap A_2 \cap \overline{A_3}) = P(A_1) \times P(A_2) \times P(\overline{A_3}) = (1 - p)^2 p;$$

etc.

7.3. Distribuição Geométrica, com parâmetro p

De uma forma geral, obtém-se que, para qualquer $k \in \mathbb{N}$,

$$P(T = k) = P(A_1 \cap A_2 \cap \dots \cap A_{k-1} \cap \overline{A_k}) = p(1 - p)^{k-1}.$$

Definição

Quando T é uma v.a. discreta, de contradomínio \mathbb{N} e função massa de probabilidade dada por

$$P(T = k) = p(1 - p)^{k-1},$$

diz-se que T segue uma distribuição Geométrica com parâmetro p , e abrevia-se por $T \sim \text{Geo}(p)$.

7.3. Distribuição Geométrica, com parâmetro p

Observações: Se $T \sim Geo(p)$,

1) Facilmente se verifica que

$$P(T > k) = (1 - p)^k,$$

para todo o $k \in \mathbb{N}$;

2) [TPC] T tem a conhecida propriedade de falta de memória

$$P(T = k + n | T > k) = P(T = n),$$

para todo o $k, n \in \mathbb{N}$;

3) O valor médio e a variância são

$$E[T] = \frac{1}{p} \quad \text{e} \quad Var[T] = \frac{1-p}{p^2},$$

respetivamente.

7.3. Distribuição Geométrica, com parâmetro p

Exemplo/Exercício: Imagine que um bêbado tem n chaves na sua carteira e que, ao chegar a casa, não consegue identificar a única chave que abre a porta. Como está tão bêbado, de cada vez que ele tenta uma chave que não é a certa, não consegue colocá-la de lado pelo que na tentativa seguinte volta a ter n chaves disponíveis para a escolha. Calcule a probabilidade de ele:

- i) acertar à primeira;
- ii) acertar pela primeira vez na terceira tentativa;
- iii) errar as primeiras 5 tentativas;
- iv) acertar pela primeira vez na oitava tentativa, sabendo que errou nas primeiras 5.

Sugestão: Identificar uma v.a. relevante para o problema e que tenha distribuição Geométrica, com parâmetro $\frac{1}{n}$.

7.4. Distribuição Uniforme no intervalo $[a, b]$

Considere uma v.a. contínua X cuja função densidade de probabilidade é constante num dado intervalo limitado $[a, b]$ e é nula fora desse intervalo. Nestas condições, a probabilidade de X assumir valores num dado sub-intervalo de $[a, b]$ será a mesma para qualquer subintervalo de igual amplitude.

Definição

Diz-se que uma v.a. contínua, X , segue a distribuição Uniforme no intervalo $[a, b]$, abrevia-se por $X \sim U([a, b])$, se a função densidade de probabilidade de X é dada por

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b \\ 0 & \text{se c.c.} \end{cases}.$$

Observe que, de facto, esta distribuição atribuí igual probabilidade a intervalos com a mesma amplitude contidos em $[a, b]$: se $]c, d[\subseteq [a, b]$ tem-se que

$$P(X \in]c, d[) = \int_c^d f(x)dx = \left[\frac{x}{b-a} \right]_c^d = \frac{d-c}{b-a} = \frac{\text{amplitude de }]c, d[}{\text{amplitude de } [a, b]}.$$

7.4. Distribuição Uniforme no intervalo $[a, b]$

Observações:

1) Se $X \sim U([a, b])$, então a função de distribuição de X é dada por [TPC]

$$F(c) = P(X \leq c) = \int_{-\infty}^c f(x)dx = \begin{cases} 0 & \text{se } c < a \\ \frac{c-a}{b-a} & \text{se } a \leq c \leq b \\ 1 & \text{se } c > b. \end{cases};$$

2) Se $X \sim U([a, b])$, então o valor médio e a variância são [TPC],

$$E[X] = \frac{a+b}{2} \quad \text{e} \quad \text{Var}[X] = \frac{(b-a)^2}{12}.$$

3) Existe ainda a distribuição Uniforme discreta que é utilizada quando se escolhe ao acaso um elemento de um conjunto finito em que os diferentes elementos têm igual probabilidade de serem escolhidos. Temos então a seguinte definição: *Seja U um subconjunto real finito, com n elementos. Diz-se que Z tem lei Uniforme no conjunto U , abrevia-se por $Z \sim \text{Uniforme}(U)$, se sua a.f.m.p. é dada por*

$$f(a) = \begin{cases} \frac{1}{n} & \text{se } a \in U \\ 0 & \text{se c.c.} \end{cases}.$$

7.5. Distribuição Exponencial, com parâmetro λ

A distribuição Exponencial é muito utilizada em estudos que envolvam tempos de espera até à ocorrência de um determinado acontecimento ou intervalos de tempo entre acontecimentos.

Definição

Diz-se que uma v.a. (absolutamente) contínua, T , segue a distribuição Exponencial com parâmetro $\lambda > 0$, abrevia-se por $T \sim \text{Exp}(\lambda)$, se a função densidade de probabilidade de T é dada por

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

7.5. Distribuição Exponencial, com parâmetro λ

Observações: Se $T \sim \text{Exp}(\lambda)$ então [TPC]:

1) a função de distribuição de T é dada por

$$F(c) = P(T \leq c) = \int_{-\infty}^c f(x)dx = \begin{cases} 0 & \text{se } c < 0 \\ 1 - e^{-\lambda c} & \text{se } c \geq 0 \end{cases} ;$$

2) T tem a conhecida propriedade de falta de memória, i.e.,

$$P(T > t + x | T > t) = P(T > x),$$

para quaisquer $t > 0, x > 0$;

3) o valor médio e a variância são

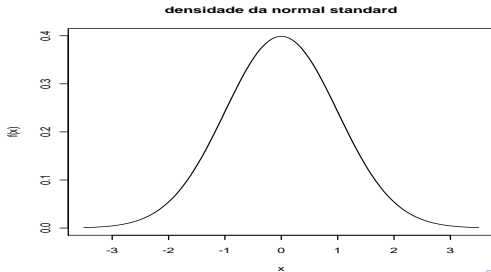
$$E[T] = \frac{1}{\lambda} \quad \text{e} \quad \text{Var}[T] = \frac{1}{\lambda^2},$$

respetivamente.

Nota: Provar 1), 2) e 3) já consta das Folhas Práticas 2 e 3. Cálculos em 3) exigem integração por partes.

7.6. Distribuição Normal, com parâmetros μ e σ^2

A distribuição Normal (ou Gaussiana) é a mais importante das distribuições contínuas, pois tem várias aplicações na modelação de situações reais e em inferência estatística. A função densidade de probabilidade desta distribuição tem uma forma de sino e é simétrica relativamente ao parâmetro μ . A título de exemplo, apresenta-se abaixo uma figura onde se pode ver a curva da função densidade de probabilidade da Normal com parâmetros 0 e 1, também conhecida por *Normal standard*.



7.6. Distribuição Normal, com parâmetros μ e σ^2

Definição

Uma v.a. contínua X segue uma distribuição Normal com parâmetros μ e σ^2 , com $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}^+$, abrevia-se por $X \sim N(\mu, \sigma^2)$, se a sua função densidade de probabilidade é dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\}, \quad x \in \mathbb{R}. \quad (3)$$

Como já foi referido atrás, no caso em que $\mu = 0$ e $\sigma = 1$, falamos em distribuição *Normal standard* ou distribuição *Normal centrada e reduzida* ou ainda distribuição *Normal padrão*.

Note-se que, qualquer que seja $a \in \mathbb{R}$, a função (3) verifica

$$f(\mu + a) = f(\mu - a)$$

mostrando-se assim que é simétrica relativamente ao parâmetro μ . Em particular, a função densidade de probabilidade da Normal standard é uma função par (i.e., é simétrica relativamente à origem).

7.6. Distribuição Normal, com parâmetros μ e σ^2

Observações/Propriedades:

1) Se $X \sim N(\mu, \sigma^2)$ então o valor médio e a variância são, respectivamente,

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

2) Se $X \sim N(\mu, \sigma^2)$ então

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

Nota: A esta operação que se efetuou à v.a. X chama-se *centrar e reduzir* a variável.

3) Se $Z \sim N(0, 1)$ então, quaisquer que sejam $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}^+$, tem-se

$$Y = \mu + \sigma Z \sim N(\mu, \sigma^2).$$

7.6. Distribuição Normal, com parâmetros μ e σ^2

Observações/Propriedades:(continuação)

4) A função de distribuição de uma $N(\mu, \sigma^2)$, naturalmente, dada por

$$c \in \mathbb{R}, F(c) = \int_{-\infty}^c \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\} dx.$$

Apesar da existência do integral, não é possível obter a expressão exata desta função. É, no entanto, possível obter aproximações numéricas! Para a distribuição Normal standard, tais aproximações constam de tabelas como a que se encontra na página seguinte.

Há que fazer uma leitura atenta das tabelas que proliferam nos livros das mais variadas áreas, pois podem representar quantidades diferentes. Em particular, na tabela que vamos utilizar (ver página seguinte), está disponível a $P(0 < Z \leq c)$ para valores de $c \in \mathbb{R}_0^+$, com $Z \sim N(0, 1)$. A simetria da função densidade de probabilidade de Z permite facilmente obter a função de distribuição de Z para qualquer outro valor de $c \in \mathbb{R}$.

Finalmente, para obter a função de distribuição de uma v.a. $X \sim N(\mu, \sigma^2)$ basta recorrer à propriedade 2) anterior, isto é, basta centrar e reduzir X .

7.6. Distribuição Normal, com parâmetros μ e σ^2

Tabela da Distribuição Normal Reduzida

Z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998
3,6	0,4998	0,4998	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,7	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,8	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999	0,4999
3,9	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000	0,5000

7.6. Distribuição Normal, com parâmetros μ e σ^2

Observações/Propriedades:(continuação)

5) Devido à simetria da função densidade de probabilidade da $N(0, 1)$ é possível concluir que a respetiva função de distribuição, aqui denotada por $F_{N(0,1)}$, satisfaz a seguinte igualdade:

$$F_{N(0,1)}(-c) = 1 - F_{N(0,1)}(c), \quad c \in \mathbb{R}.$$

Em particular, isto implica que é válida a seguinte relação entre o quantil de ordem p e o quantil de ordem $1 - p$ da distribuição Normal standard: para todo o $0 < p < 1$,

$$\chi_p = -\chi_{1-p}.$$

6) A tabela aqui utilizada e a simetria da função densidade de probabilidade de uma v.a. $X \sim N(0, 1)$ permitem facilmente calcular $P(|X| \leq b)$, $b \in \mathbb{R}^+$. Note que, neste caso,

$$P(|X| \leq b) = 2P(0 < Z \leq b).$$

Este tipo de cálculos surge frequentemente em problemas de inferência estatística, nomeadamente na dedução de intervalos de confiança e de testes de hipóteses paramétricos.